Emerging Topics in Statistics and Biostatistics

Ding-Geng (Din) Chen
Samuel O. M. Manda
Tobias F. Chirwa *Editors*

Modern Biostatistical Methods for Evidence-Based Global Health Research

Springer

Emerging Topics in Statistics and Biostatistics

Ding-Geng (Din) Chen • Samuel O. M. Manda •
Tobias F. Chirwa
Editors

Modern Biostatistical Methods for Evidence-Based Global Health Research

Springer

Editors
Ding-Geng (Din) Chen
College of Health Solutions
Arizona State University
Phoenix, AZ, USA

Samuel O. M. Manda
Department of Statistics
University of Pretoria
Pretoria, South Africa

Tobias F. Chirwa
School of Public Health, Division of
Epidemiology and Biostatistics
University of the Witwatersrand
Johannesburg, South Africa

ISSN 2524-7735 ISSN 2524-7743 (electronic)
Emerging Topics in Statistics and Biostatistics
ISBN 978-3-031-11011-5 ISBN 978-3-031-11012-2 (eBook)
https://doi.org/10.1007/978-3-031-11012-2

Preface

The DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) training program is funded by the Wellcome Trust in partnership with the Alliance for Accelerating Excellence in Science in Africa (AESA). The consortium was established in 2015, with the overall aim of building a critical mass of biostatisticians and biostatistics research leadership in Sub-Saharan Africa. This is achieved through the development and strengthening of biostatistics capacity and resource at the 11 participating local institutions, in collaboration with four local research institutions and three northern university partners.

In celebrating the contributions, achievements, and progress of SSACAB scientists and their partners and collaborators, this book is organized to document the contributions from the consortium with a diverse mix of current scholarship and exposition of biostatistics methods and application for evidence-based global health in the Region. The volume features inspiring and informative chapters that reflect on the accomplishments of biostatistics research and its applications that offer solutions to local health problems. There are a total of 18 chapters to provide an overview of the emerging topics in biostatistical methods and their applications to Sub-Saharan Africa public health research and evidence-based management decision-making.

The structure of these 18 chapters is subsequently organized with the following five parts. As an introductory chapter, chapter "Sub-Saharan African Region Strategies to Improve Biostatistics Capacity: Exploring Collaborations Between Training and Research Institutions," describes the origins and contributions of SACCAB as well as its structure.

Part I (Data Harmonization and Analysis) contains three chapters (Chapters 2 to 4). In chapter "Diagonal Reference Modelling of the Effects of Educational Differences Between Couples on Women's Health-Care Utilization in Eritrea," Ghilagaber adapted models developed in the social mobility literature to examine the effects of differences between couples' educational levels on women's health-related decisions (such as the propensity to deliver in health facilities). Both conventional modeling and Diagonal Reference Modeling (DRM) which account for origin (woman's education), destination (partner's education), and "mobility" (differences between couples' educational levels) are applied on data from Demo-

graphic and Health Surveys (DHS). Results from conventional models reveal strong effects of educational differences on women's health-related decisions, but such strong effects disappear when data is analyzed using DRM. In chapter "Sequential Probit Modeling of Regional Differences in the Effects of Education on Parity Progression Ratios in Ethiopia," Ghilagaber and Peristera proposed a sequential procedure to model differentials in parity progression in Ethiopia based on data from its 2019 Mini Demographic and Health Survey in which 8885 women from 11 regions were interviewed. Their results showed that the sequential model provides more insight than conventional models when exploring the association between education and parity progression in particular and fertility decision process in general. They also found both similarities and differences in the effects of education on parity progression among the regions. In chapter "Propensity Score Approaches for Estimating Causal Effects of Exposures in Observational Studies," Twabi and Manda assessed causal effects of maternal health (including HIV infection) and breastfeeding practices on child health outcomes. They offered a statistical causal inference method to rigorously investigate the purported causal relationships of maternal HIV infection, nutritional status, and breastfeeding practices on child health outcomes from population-based nationally representative data from Demographic and Health Surveys in Malawi and Zambia.

Part II (Systematic Review and Statistical Meta-Analysis) is organized with four chapters (Chapters 5 to 8). In chapter "Evidence-Informed Public Health, Systematic Reviews and Meta-analysis," Abariga, Ayele, McCaul, Musekiwa, Ochodo, and Rohwer used systematic reviews, statistical meta-analysis, and illustrative examples relevant to Sub-Saharan Africa that can be used to inform public health decisions. They unpacked aspects that need to be considered when performing meta-analysis including statistical tests to use, assessment of heterogeneity, subgroup analysis, meta-regression, and sensitivity analysis. Furthermore, they covered emerging techniques in the meta-analysis, including network meta-analysis, multivariate meta-analysis, data synthesis when meta-analysis is not possible, and meta-analysis of diagnostic test accuracy (DTA) studies. In chapter "Statistical Meta-analysis and Its Efficiency: A Real Data Analysis and a Monte-Carlo Simulation Study," Chen gave an overview of meta-analysis on classical fixed-effects and random-effects to synthesize summary statistics as well as meta-regression to explain the between-study heterogeneity. A Monte-Carlo simulation study was designed to illustrate the relative efficiency of the MA using summary statistics to the MA using the original individual participant-level data. Real meta-data from 13 clinical trials to assess the Bacillus Calmette-Guerin vaccine in the prevention of tuberculosis was used to demonstrate the implementation of these meta-analysis models. In chapter "Meta-Analysis Using R Statistical Software," Onyango and Wao introduced a series of topics in systematic review and meta-analysis (SRMA). They used illustrative examples to demonstrate how SRMA is undertaken for one continuous and one dichotomous outcome. In chapter "Longitudinal Meta-analysis of Multiple Effect Sizes," Musekiwa and colleagues discussed the meta-analysis from multiple outcomes where multiple effect sizes are estimated and produced. These estimated effect sizes could be correlated because they are measured from

the same studies. Additionally, the outcomes are often measured longitudinally, resulting in multiple effect sizes estimated repeatedly over time. This chapter proposes methods for statistical meta-analysis combining summary data from more than one longitudinal study with multiple effect sizes. The proposed methods were illustrated by an analysis of an example involving longitudinal meta-analysis of HIV studies assessing the effect of some antiretroviral drugs in improving viral load suppression and increasing CD4 count at weeks 4, 8, 12, 16, 20, 24, 32, 40, and 48 after start of treatment assignment.

Part III (Spatial-Temporal Modelling and Disease Mapping) consists of two chapters (Chapters 9 to 10). In chapter "Measuring Bivariate Spatial Clustering in Disease Risks," Darikwa and Manda compared a set of full Bayesian estimations for fitting a multivariate spatial disease model. They applied the models to age-gender all-cause mortality in South Africa and childhood illnesses in Malawi. The effect on the degree of spatial correlation after adjusting for socio-demographic factors previously associated with studies diseases is also assessed. In chapter "Bivariate Copula-Based Spatial Modelling of Health Care Utilisation in Malawi," Gondwe, Chipeta, and Kazembe constructed three joint models: first to analyze the distribution of mixed binary-continuous data, a second for a mixture of a count and continuous variables, and a third for a discrete set of count and binary variables. The models are applied to study ANC utilization among Malawian women using the 2015 Malawi Demographic and Health Survey (MDHS) data, drawn using a stratified cross-sectional survey design. The models allowed for simultaneous estimation of dependence and marginal distribution parameters of timing and frequency of healthcare utilization to understand factors influencing utilization. Covariates included demographics, socio-economic factors, and location. Various models were fitted and compared, assuming different spatial structures.

Part IV (Bayesian Statistical Modelling) is composed of four chapters (Chapters 11–14). In chapter "Bayesian Survival Analysis with the Extended Generalized Gamma Model: Application to Demographic and Health Survey Data," Liang and Ghilagaber extended the existing family of flexible survival models by assembling models scattered across the literature into a more knit-form and under the same umbrella. New special cases are obtained not only by constraining the shape and scale parameters of the extended generalized gamma (EGG) model to fixed constants but also by imposing relationships (such as reciprocal) between them. The models were illustrated using data on family initiation from Demographic and Health Surveys in some Sub-Saharan African countries. Preliminary results showed that the further extended family of distributions provided a wide range of alternatives for a baseline distribution in the analysis of survival data. In chapter "Dynamic Bayesian Modeling of Educational and Residential Differences in Family Initiation Among Eritrean Men and Women," Munezero and Ghilagaber proposed a dynamic Bayesian survival model in analyzing differentials in the timing of family initiation. Such formulation relaxed the strong assumption of constant hazard ratio in conventional proportional hazard models and allows covariate effects to vary over time. The inference is fully Bayesian and efficient sequential Monte Carlo (Particle Filter) is used to sample from the posterior distribution. They illustrated

the proposed model with data on entry into first marriage among Eritrean men and women surveyed in the 2010 Eritrean Population and Health Survey. Results from the conventional proportional hazards model indicated significant differences in family initiation among all educational and residential groups. In the dynamic model, on the other hand, only one educational and one residential group among the women and only one residential group among the men differed from their respective baseline groups. In chapter "Bayesian Spatial Modeling of HIV Using Conditional Autoregressive Model," Ogunsakin and Chen proposed a generalized linear model (GLM) with Bayesian inference to build the Spatially Varying Coefficients model and compared it with the stationary model to evaluate the spatial association between the incidence of HIV and some socio-demographic risk factors in Nigeria. They found a nonlinear relationship between the incidence of HIV and age. The modeling of the socio-demographic predictors of HIV infection and spatial maps provided in this study could aid in developing a framework to alleviate HIV and identify its hotspots for urgent intervention in the endemic regions. In chapter "Estimating Determinants of Stage at Diagnosis of Breast Cancer Prevalence in Western Nigeria Using Bayesian Logistic Regression," Ogunsakin and Chen estimated the prevalence and investigated determinants of stage at diagnosis by constructing Bayesian logistic regression model from a generalized linear modeling using socio-economic, demographic, and medical factors. They established that age, higher educational level, being a westerner, as well as choosing nursing as a career were the major factors that motivate early stage at breast cancer diagnosis in this part of Nigeria and that delays in diagnosis reflect a lack of education. They recommend an intensive health education program in order to increase early-stage diagnosis for patients.

Part V (Statistical Applications) has four chapters (Chapters 15–18) to discuss the statistical methods and applications in longitudinal data, survival data, and missing data imputation. In chapter "Identifying Outlying and Influential Clusters in Multivariate Survival Data Models," Kaombe and Manda developed methods for group outlier and influence assessments for the time-independent clustered survival model. Appropriate extensions of martingale-based residuals in univariate survival model and the re-weighted minimum covariance determinant method in multivariate linear mixed-effects model have been defined for group outlier analysis for the clustered survival model. They adapted influence approximations based on the one-step Newton-Raphson method for maximum likelihood estimators in univariate survival analysis to develop a group influence method for the survival mixed model. They demonstrated the performance of the proposed methods through a simulation study and real data application. In chapter "Joint Modelling of Longitudinal and Competing Risks Survival Data," Masangwi, Muula, and Mukaka used a joint modeling framework to combine the three blocks in the analysis. The methods were applied to the malaria dataset from Malawi where longitudinal markers hemoglobin level and parasite count were considered. Time to treatment failure due to severe malaria and time to withdrawal were the survival outcomes. Different survival outcomes were observed, and they noted that when there is an association between longitudinal and survival outcomes in biomedical research, joint models should be

considered as they performed better than the separate methods. But where there is no association, separate models for survival and longitudinal data analysis should be considered. In chapter "Stratified Multilevel Modelling of Survival Data: Application to Modelling Regional Differences in Transition to Parenthood in Ethiopia," Ghilagaber, Akinyi Lagehäll, and Yemane presented a multilevel extension of the Cox proportional hazards model where a shared frailty term is included to account for clustering of women within households. The extended model is used to analyze regional differences in the intensity of transition to parenthood among 15,019 Ethiopian women aged 15–49 years old in the country's Demographic and Health Survey of 2016. They found that household frailty effects are fairly small in the nine regions, but the log-normal frailties were significant in the entire country and the two city administrations which are relatively heterogeneous with inhabitants from many ethnic groups. They also found regional differences in the effects of the background variables on the intensity of transition to parenthood, but the effects were generally stable across the three models in each region. In chapter "Application of Multiple Imputation, Inverse Probability Weighting, and Double Robustness in Determining Blood Donor Deferral Characteristics in Malawi," Kudowa, Mavuto, and Mukaka addressed missing data to a retrospective cohort involving blood donor data to estimate predictors of donor deferral status. The logistic regression model was fit on deferral status and the independent variables. Multiple Imputation by Chained Equation, Inverse Probability Weighting (IPW), and Double Robustness (DR-IPW) were applied to correct for the missingness. The estimates from these methods were compared with estimates from the CC method.

We sincerely thank all of the people who have given us strong support for the publication of this book on time. Our acknowledgments go to all the chapter authors (in the "List of Contributors") for submitting the excellent works to this book. We deeply appreciate the reviews of many reviewers (in the "List of Chapter Reviewers"). Their comments and suggestions have improved the quality and presentation of the book substantially. Last but not least, we are so grateful to Laura Aileen Briskman and Eva Hiripi (Editors, Statistics, Springer Nature) and Kirthika Selvaraju (Project Coordinator of Books, Springer Nature) for their full support during the long publication process. We look forward to receiving comments about the book from the readers. For any suggestions about further improvements to the book, please contact us by email.

Phoenix, AZ, USA — Ding-Geng (Din) Chen
Pretoria, South Africa — Samuel O. M. Manda
Johannesburg, South Africa — Tobias F. Chirwa

Contents

Sub-Saharan African Region Strategies to Improve Biostatistics Capacity: Exploring Collaborations Between Training and Research Institutions 1
Tobias F. Chirwa, Pascalia O. Munyewende, Ding-Geng (Din) Chen, and Samuel O. M. Manda

Part I Data Harmonization and Analysis

Diagonal Reference Modelling of the Effects of Educational Differences Between Couples on Women's Health-Care Utilization in Eritrea 9
Gebrenegus Ghilagaber

Sequential Probit Modeling of Regional Differences in the Effects of Education on Parity Progression Ratios in Ethiopia 21
Gebrenegus Ghilagaber and Paraskevi Peristera

Propensity Score Approaches for Estimating Causal Effects of Exposures in Observational Studies 41
Halima S. Twabi, Samuel O. M. Manda, and Dylan S. Small

Part II Systematic Review and Statistical Meta-Analysis

Evidence-Informed Public Health, Systematic Reviews and Meta-Analysis 89
Samuel A. Abariga, Michael McCaul, Alfred Musekiwa, Eleanor Ochodo, and Anke Rohwer

Statistical Meta-Analysis and Its Efficiency: A Real Data Analysis and a Monte-Carlo Simulation Study 137
Ding-Geng (Din) Chen

Meta-Analysis Using R Statistical Software 167
Nelson Owuor Onyango and Hesborn Otieno Wao

Longitudinal Meta-Analysis of Multiple Effect Sizes 203
Alfred Musekiwa, Samuel O. M. Manda, Henry G. Mwambi, Ding-Geng (Din) Chen, Samuel A. Abariga, Michael McCaul, Eleanor Ochodo, and Anke Rohwer

Part III Spatial-Temporal Modelling and Disease Mapping

Measuring Bivariate Spatial Clustering in Disease Risks 235
Timotheus B. Darikwa and Samuel O. M. Manda

Bivariate Copula-Based Spatial Modelling of Health Care Utilisation in Malawi 261
Ellen Gondwe, Michael G. Chipeta, and Lawrence Kazembe

Part IV Bayesian Statistical Modelling

Bayesian Survival Analysis with the Extended Generalized Gamma Model: Application to Demographic and Health Survey Data ... 287
Yuli Liang and Gebrenegus Ghilagaber

Dynamic Bayesian Modeling of Educational and Residential Differences in Family Initiation Among Eritrean Men and Women 319
Parfait Munezero and Gebrenegus Ghilagaber

Bayesian Spatial Modeling of HIV Using Conditional Autoregressive Model 339
Ropo Ebenezer Ogunsakin and Ding-Geng (Din) Chen

Estimating Determinants of Stage at Diagnosis of Breast Cancer Prevalence in Western Nigeria Using Bayesian Logistic Regression 355
Ropo Ebenezer Ogunsakin and Ding-Geng (Din) Chen

Part V Statistical Applications

Identifying Outlying and Influential Clusters in Multivariate Survival Data Models 377
Tsirizani M. Kaombe and Samuel O. M. Manda

Joint Modelling of Longitudinal and Competing Risks Survival Data 411
Didjier D. Masangwi, Adamson S. Muula, and Mavuto F. Mukaka

Stratified Multilevel Modelling of Survival Data: Application to Modelling Regional Differences in Transition to Parenthood in Ethiopia 431
Gebrenegus Ghilagaber, Amanda Akinyi Lagehäll, and Elelta Yemane

Application of Multiple Imputation, Inverse Probability Weighting, and Double Robustness in Determining Blood Donor Deferral Characteristics in Malawi 457
Evaristar N. Kudowa and Mavuto F. Mukaka

Index 475

Contributors

Samuel A. Abariga Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Ding-Geng (Din) Chen College of Health Solutions, Arizona State University, Phoenix, AZ, USA
Department of Statistics, University of Pretoria, Pretoria, South Africa

Michael G. Chipeta University of Oxford, Big Data Institute, Oxford, UK

Tobias F. Chirwa School of Public Health at University of the Witwatersrand, Johannesburg, South Africa

Timotheus B. Darikwa Department of Statistics and Operations Research, University of Limpopo, Sovenga, South Africa

Gebrenegus Ghilagaber Department of Statistics, Stockholm University, Stockholm, Sweden

Ellen Gondwe University of Malawi, Chancellor College, Zomba, Malawi

Tsirizani M. Kaombe Department of Mathematical Sciences, Chancellor College, University of Malawi, Zomba, Malawi

Lawrence Kazembe University of Namibia, Windhoek, Namibia

Evaristar N. Kudowa University of Malawi, Chancellor College, University of North Carolina Project-Malawi, Lilongwe, Malawi

Amanda Akinyi Lagehäll Department of Statistics, Stockholm University, Stockholm, Sweden

Yuli Liang School of Business and Economics, Department of Economics and Statistics, Linnaeus University, Växjö, Sweden

Samuel O. M. Manda Department of Statistics, University of Pretoria, Pretoria, South Africa
Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa

Danger D. Masangwi University of Malawi, Chancellor College, Zomba, Malawi
Malawi University of Science and Technology, Mikolongwe, Malawi

Michael McCaul Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Mavuto F. Mukaka Malawi University of Science and Technology, Mikolongwe, Malawi
Nuffield Department of Medicine, University of Oxford, Oxford, UK
Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Nakhon Pathom, Thailand

Parfait Munezero Department of Statistics, Stockholm University, Stockholm, Sweden

Pascalia O. Munyewende School of Public Health at University of the Witwatersrand, Johannesburg, South Africa

Alfred Musekiwa Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Adamson S. Muula University of Malawi, College of Medicine, Zomba, Malawi
Africa Center of Excellence in Public Health and Herbal Medicine, University of Malawi, Zomba, Malawi

Henry G. Mwambi University of KwaZulu-Natal, Scottsville, South Africa

Eleanor Ochodo Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Ropo Ebenezer Ogunsakin Department of Statistics, University of Pretoria, Pretoria, South Africa
Kapenda S.R.N University of Namibia, Department of Statistics and Population Studies, Windhoek, Namibia

Nelson Owuor Onyango School of Mathematics, University of Nairobi, Nairobi, Kenya

Paraskevi Peristera Stockholm University, Stress Research Institute, Department of Psychology, Stockholm, Sweden

Anke Rohwer Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Dylan S. Small Department of Statistics, University of Pennsylvania, Philadelphia, PA, USA

Halima S. Twabi Department of Mathematical Sciences, University of Malawi, Zomba, Malawi

Hesborn Otieno Wao African Population and Health Research Center, Nairobi, Kenya

Elelta Yemane Department of Statistics, Stockholm University, Stockholm, Sweden

List of Chapter Reviewers

Yusentha Balakrishna South African Medical Research Council, Cape Town, South Africa

Michael Chipeta Department of Crop and Soil Sciences, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi

Timotheus B. Darikwa Department of Statistics and Operations Research, University of Limpopo, Polokwane, South Africa

Romain Lucas Glèlè Kakaï Biostatistics and forest estimations, Faculty of Agronomic Sciences (FSA in French), University of Abomey-Calavi, Godomey, Benin

Ngianga-Bakwin Kandala WMS - Population Evidence and Technologies, Warwick University, Coventry, UK

Lawrence Kazembe University of Namibia, Windhoek, Namibia

Thomas A. Louis Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Samuel O. M. Manda Biostatistics Research Unit, South Africa Medical Research Council, Pretoria, South Africa

Innocent Maposa Epidemiology & Biostatistics Division, University of the Witwatersrand, Johannesburg, South Africa

Mavuto F. Mukaka Malawi University of Science and Technology, Mikolongwe, Malawi
Nuffield Department of Medicine, University of Oxford, Oxford, UK
Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Nakhon Pathom, Thailand

Alfred Musekiwa Epidemiology and Biostatistics Department, School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa

Henry G. Mwambi School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Justine Nasejje School of Statistics, University of the Witwatersrand, Johannesburg, South Africa

Nelson Owuor Onyango School of Mathematics, University of Nairobi, Nairobi, Kenya

Tarylee Reddy South African Medical Research Council, Durban, KwaZulu-Natal, South Africa

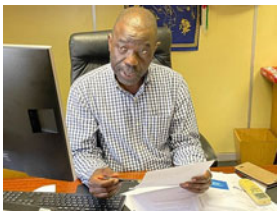
Jim Todd London School of Hygiene and Tropical Medicine, TAZAMA Project NIMR, Mwanza, Tanzania

Patrick Guge Weke School of Mathematics, University of Nairobi, Nairobi, Kenya

About the Editors

Ding-Geng (Din) Chen is Elected Fellow of the American Statistical Association (ASA) and Elected Member of the International Statistics Institute (ISI). He is currently the Executive Director and Professor in Biostatistics at the College of Health Solutions, Arizona State University. He is also the South Africa Research Chair Initiatives (SARChI) Research Chair in Biostatistics (Tier 1) established by the Department of Science and Technology (DST), National Research Foundation (NRF), and South Africa Medical Research Council (SAMRC), to lead the research, development, and capacity building in biostatistics across South Africa, an extraordinary professor in the Department of Statistics, University of Pretoria, South Africa, and Honorary Professor at the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa. Professor Chen served as distinguished Professor in Biostatistics at the University of North Carolina-Chapel Hill, Biostatistics Professor at the University of Rochester Medical Center, and the Karl E. Peace Endowed Eminent Scholar Chair in Biostatistics at the Jiann-Ping Hsu College of Public Health at the Georgia Southern University. Professor Chen has written more than 200 refereed publications and co-authored/co-edited 33 books on clinical trial methodology, meta-analysis, causal inference, and public health statistics. This work is partially supported by the National Research Foundation of South Africa (Grant Number 127727) and the South African National Research Foundation (NRF) and South

African Medical Research Council (SAMRC) (South African DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant Number 114613).



Samuel O. M. Manda is the Head and Professor of Statistics in the Department of Statistics, University of Pretoria, Pretoria, South Africa. He holds a PhD in Bayesian Statistics from the University of Waikato, and subsequently, he did postdoctoral studies in nonparametric Bayesian survival statistics at the University of Auckland, New Zealand. He has served as Director of the Biostatistics Research Unit at the South African Medical Research Council, South Africa. Previously, Prof Manda was at the University of Leeds, United Kingdom. Prof Manda has over 130 refereed publications on spatial statistics, survival data, meta-analyses, and causal inference. He is a National Research Foundation rated scientist and has honorary appointments including Extraordinary Professorships (Stellenbosch University and North-West University, South Africa), KwaZulu-Natal (Honorary Professor), and Malawi (Adjunct Professor). Prof Manda's science contributions also cover postgraduate supervision and membership of national and international technical committees. Over the years, he has supported several national health projects including health surveys and clinical trials as a lead statistic. He has also supported several biostatistics capacity development activities, mainly through the DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB).

Tobias F. Chirwa is the Program Director for the Sub-Saharan African Consortium for Advanced Biostatistics (SSACAB) training program. He is the Head of the School of Public Health of the University of the Witwatersrand, Johannesburg (Wits). Prior to this, he was the Head of Division of the Epidemiology and Biostatistics and managed the Wits Faculty of Health Sciences statistical support services. He was Senior Lecturer at the University of Malawi (UNIMA) and Research Fellow in Medical Statistics at the London School of Hygiene and Tropical Medicine (LSHTM). He has also held positions of responsibility like the Dean, Faculty of Science, and Head, Department of

Mathematical Sciences at UNIMA and was also the President of Statistical Association of Malawi. He has led postgraduate capacity building initiatives including the Norwegian-funded NUFU PhD program, MSc Biostatistics and Informatics at UNIMA. Further, he has continued to be active in curriculum development while at Wits – developing short courses in biostatistics; MSc programs in Biostatistics and Research Data Management. He is also Co-Director of a University of North Carolina (UNC) and Wits Fogarty (MSc and PhD) training grant on AIDS Implementation Science and Advanced Cohort Analyses, together with the WHO/TDR Implementation Research.

Sub-Saharan African Region Strategies to Improve Biostatistics Capacity: Exploring Collaborations Between Training and Research Institutions

Tobias F. Chirwa, Pascalia O. Munyewende, Ding-Geng (Din) Chen, and Samuel O. M. Manda

Abstract There has been an increase in health sciences research conducted within the sub-Saharan African (SSA) region in connection with the rest of the world. However, the capacity to analyse the generated data to support public health policies has been limited. Several initiatives aimed at building and retaining biostatistics resources and capacity in the region have been implemented with differing successes, scope and coverage. One such initiative is the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA), the DELTAS Africa Sub-Saharan African Consortium for Advanced Biostatistics (SSACAB) training programme. The DELTAS Africa SSACAB training programme was created to address the dearth of biostatistical capacity in the SSA region. It relies on the principle of pooling together limited biostatistics capacity in the African region to increase the numbers of trained fellows through collaborative masters and doctoral training. This book showcases some of the research work that has been undertaken under SACCAB. In this introductory chapter, we describe the origins and contributions of SACCAB as well as its structure. A total of 150 fellows of which 123 are masters fellowships (41 female) have been produced under SACCAB.

Keywords Biostatistics · Capacity building · DELTAS Africa · SSACAB · Programme achievements · Networks and partnerships · Sub-Saharan Africa

T. F. Chirwa (✉) · P. O. Munyewende
School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
e-mail: tobias.chirwa@wits.ac.za

D.-G. (Din) Chen
College of Health Solutions, Arizona State University, Phoenix, AZ, USA

Department of Statistics, University of Pretoria, Pretoria, South Africa

S. O. M. Manda
Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_1

1 Introduction

Health sciences research plays a key role in strengthening health systems, providing evidence-based interventions which help to inform policy and practice (Agnandji et al., 2012; Franzen et al., 2017). Although a lot of data has been generated from such initiatives and largely driven by local research institutions in most sub-Saharan African (SSA) countries (Gezmu et al., 2011), there is limited statistical and biostatistical capacity to analyse such data (Thomson et al., 2016). The limited biostatistical capacity often based in universities and research institutions is often overstretched (Gezmu et al., 2011; Machekano et al., 2015).

Training biostatisticians abroad is not cost-effective compared to utilising existing local institutions. Further, many biostatisticians trained abroad rarely return to their home countries. A group of local biostatisticians dotted within the SSA region took the opportunity and started an initiative to build the critical mass to fill the urgent need for biostatisticians in the region. The group of biostatisticians piggybacked on the existing limited post-graduate training programmes and research institutions that provided cutting-edge health sciences research questions and learning experiences for masters and doctoral fellows (Machekano et al., 2016; Thomson et al., 2016). This resulted in the formation of the DELTAS Africa SSACAB training programme (Chirwa et al., 2020). The DELTAS Africa SSACAB programme constitutes 11 African universities in 9 countries, 4 research institutions and 4 Northern partners led by the School of Public Health based at the University of the Witwatersrand, Johannesburg. Chirwa et al. (2020) highlight and describe the eight other initiatives in the region which are geared towards developing biostatistics capacity. SSACAB is based on the principle of pooling limited biostatistics resources in training and research institutions to teach and supervise postgraduate students at various partner universities.

One of the research institutions which have actively contributed to the success of SSACAB is the South Africa Medical Research Council (SA MRC) Biostatistics Unit. The biostatistics unit is one of the intra-mural capacity building efforts based in the SAMRC whose mandate is to provide biostatistics expertise and support to the organisation's network of medical and health researchers as well as government departments and national and international research bodies and is an interdisciplinary unit with expertise in biostatistics, GIS, data management and food science. The k*ey focus areas for the* biostatistics unit are to develop generic, innovative and rigorous statistical methodology that improves the design and analysis of health studies and to provide biostatistical leadership and expertise in collaborative health and medical-related research projects. The unit also produces and uses health-related maps and dietary intake research tools in the national food database. The SAMRC Biostatistics Unit is also responsible for ensuring that a high standard of data quality is achieved in all studies through rigorous data management by providing relevant biostatistical knowledge and support to inform local, national and international public health policy through service on projects

and review boards. The unit supports SSACAB capacity-building efforts through postgraduate supervision and specialised training in South Africa and in the region.

More recently, we have seen growth, especially in South Africa, to support capacity for advanced biostatistics through research chairs. One such unique chair, based at the University of Pretoria, was funded by the South African Research Chair Initiative (SARChI) in 2018 and supported by the Department of Science and Technology (DST), the National Research Foundation (NRF) and the South African Medical Research Council (SAMRC). There are several long-term objectives for this SARChI Research Chair in Biostatistics. Firstly, this SARChI Research Chair in Biostatistics is to develop novel biostatistics methodologies for designing and building appropriate foundations for health research and interventions that will provide the tools for building classical and adaptive research interventions to meet cost-effective public health needs and interventions in South Africa. Secondly, this SARChI Research Chair in Biostatistics is to develop novel methodologies for analysing rich and complex biostatistical data obtained from intensive longitudinal research in public health, cancer epidemiology, bioinformatics and genetics, HIV/AIDS and malaria intervention and management. Thirdly, the SARChI Research Chair in Biostatistics is to develop computational software and tools for use by researchers in public health and to publicise these novel methodologies to facilitate their application and implementation. Some of these software and tools can be patented, and some will be distributed free for immediate public use and to train faculty members, students and public health researchers by means of short courses and seminars during national/international conferences and advanced courses in biostatistics to train qualified biostatisticians needed to address the copious enormous number of public health issues facing SA today. Since 2018, the SARCHI Research Chair in Biostatistics has published 12 books in biostatistics and public health in internationally known journals. Forty-eight referred papers are also featured in public available journals highlighting the importance of biostatistical methods and applications for solving public health problems. The SARChI Research Chair in Biostatistics has also presented six keynote presentations at international conferences and has taught eight lectures at international biostatistics workshops. In terms of building capacity, the SARChI Research Chair in Biostatistics has trained and mentored 5 postdocs, 3 PhD students, 8 masters fellows and 26 honours students.

2 SSACAB

A total of 150 (masters and PhD) fellows have been awarded scholarships to date from 14 different countries in SSA. Since the inception of the SSACAB in 2016, a total of 123 masters have been awarded a fellowship as of 2018. Of these 90% have completed their MSc degrees in biostatistics and graduated. Thirteen masters graduates have been enrolled in PhD programmes either in the same institutions that they graduated from such as KCMCo or in partner institutions two at KWTRP and

collaborating programmes, Malawi Liverpool Wellcome Trust Programme, Catholic University of Health and Allied Sciences in Tanzania and in the universities of Cape Town and Kwa-Zulu Natal as well as three who are in the United Kingdom and other parts of SSA highlighting SSACAB's global reach and clear career pathing goals.

To date, our masters and PhD fellows have been able to publish more than 60 research articles in peer-reviewed journals such as Frontiers, BioMed Central, journals of medical statistics and informatics, Lancet Global Health, Geospatial Health, Statistical Methods in Medical Research, Research in Mathematics & Statistics and many other high-impact journals. Their publications address various research areas using longitudinal data analysis (malaria, HIV repeated measures), machine learning, spatial analysis (malaria clustering, malnutrition, mortality), transition modelling (HIV staging, family formation and dissolution) and stochastic and deterministic modelling (nutrition interventions) to address regional public health challenges.

Although initially SSACAB had planned 15 PhD fellowships, a total of 27 (10 female) PhD students have been offered fellowships. Of these, six PhDs are now pursuing post-doctoral fellowships or working as lecturers in various African countries.

There are currently more than 40 peer-reviewed publications from PhD fellows. SSACAB has also partially supported other PhD students enrolled in partner institutions with their manuscript publication fees in peer-reviewed open-access journals. Furthermore, staff members within SSACAB have also been supported in publishing their research work, while some have presented their work at international conferences such as The Sub-Saharan Africa Network (SUSAN) of the International Biometrics Society (IBS), the South African Statistical Association and conferences organised by SSACAB since 2017 to 2021. Partially supported staff and student research have resulted in approximately 65 peer-reviewed and open-access publications. Some staff-supported research has resulted in the publication of books, including the Statistical Modelling of Complex correlated and clustered data using household surveys in Africa edited volume from the University of Namibia (Ngianga-Bakwin & Lawrence, 2019).

3 Collaborations with Training Institutions

The SAMRC BSU collaborated with the MASAMU Program at Auburn University (funded by the National Science Foundation (NSF)). One of its objectives is to enhance research in the mathematical sciences within Southern Africa Mathematical Sciences Association (SAMSA) institutions and the African Institute for Mathematical Sciences (AIMS). The AIMS has six centres of excellence across Africa, in Ghana, Cameroon, Senegal, Tanzania and Rwanda, and South Africa. AIMS's objective is to enable Africa's talented students to become innovators driving the continent's scientific, educational and economic self-sufficiency. The

South African Centre for Epidemiological Modelling and Analysis (SACEMA), another collaborator, is a national research centre dedicated to modelling and analysis to improve health in South Africa and across the African continent. SACEMA offers training in mathematics, biology, physics, economics, statistics and epidemiology; we bridge disciplines to understand disease dynamics and improve real-world outcomes. The Wellcome Trust African Institutions Initiative (AII) through several consortia (e.g. SSACAB, CARTA and other regional initiatives including Training Health Researchers into Vocational Excellence in East Africa (THRiVE)) links academic and research institutions from Uganda, Rwanda, Tanzania and Kenya. The Netherlands–African Partnership for Capacity Development and Clinical Interventions of Poverty-related Diseases (NACCAP), which builds research capacity between several sub-Saharan African academic institutions with support from Dutch partners; the Health Research Capacity Strengthening Initiative partnership between the UK Department for International Development (DFID), the International Development and Research Centre (IDRC) Canada and the Wellcome Trust, SACORE and BAPED, have made significant progress to build research capacity.

These initiatives have had various degrees of success. However, most have been rather disjointed and more focused on HIV/AIDS, TB, child and maternal health with statisticians leaving to private industry, as there are few academic centres for biostatistics that are tightly linked to local biomedical research. Perhaps, one of the most concerning issues has been a lack of systematic and rigorous interrogation of the data being used. South Africa and the sub-Saharan African (SSA) region generate huge amounts of health data from a variety of sources including demographic and health surveillance sites (DHSS), regional and nationally representative health surveys and Routine Health Information Systems (RHIS). These data have varying concerns regarding completeness, timeliness, representativeness and accuracy. However, their utilisation remains sub-optimal because optimal analyses of such data demand an in-depth assessment and investigation of data and the process and design that generated it. Moreover, current postgraduate training in biostatistics has tended to produce ‘data analysts’, with heavy reliance on implementation of developed biostatistics techniques in the widely available statistical software. Seldom have this training embedded development and validation of methods relevant to the problem at hand. SACCAB tried to blend the two: reliance on implementation of developed biostatistics techniques in the widely available and development and validation of methods relevant.

4 Conclusion

There has been tremendous progress in terms of capacity building and research for students as part of their training and learning experiences. While others have had opportunities to publish their work, some were limited and could not showcase their work and advanced biostatistical skills gained through research and analysis

conducted. This book, therefore, provides an excellent opportunity to the readers to see the high-level analysis conducted over the 5-year period or so by locally trained biostatisticians to answer cutting-edge research questions within the SSA region. It not only aims to show evidence-based decisions based on such analysis but also that, if the right candidates are identified, nurtured and mentored, the region is able to support research without exporting data for analysis abroad.

Acknowledgements This work was supported through the DELTAS Africa Initiative Grant No. 107754/Z/15/Z-DELTAS Africa SSACAB. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

This work was also supported by the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics to Professor Ding-Geng (Din) Chen, Grant number 114613). Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF and SAMRC.

References

Agnandji, S. T., et al. (2012). Patterns of biomedical science production in a sub-Saharan research center. *BMC Medical Ethics, 13*(3).

Chirwa, T. F., Matsena Zingoni, Z., Munyewende, P., et al. (2020). Developing excellence in biostatistics leadership, training and science in Africa: How the Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) training unites expertise to deliver excellence [version 2; peer review: 2 approved, 1 approved with reservations]. *AAS Open Res, 3*, 51. https://doi.org/10.12688/aasopenres.13144.2

Franzen, S. R. P., Chandler, C., & Lang, T. (2017). Health research capacity development in low and middle income countries: Reality or rhetoric? A systematic meta-narrative review of the qualitative literature. *BMJ Open, 7*, e012332. https://doi.org/10.1136/bmjopen-2016-012332

Gezmu, M., et al. (2011). Strengthening biostatistics resources in Sub-Saharan Africa: Research Collaborations through U.S. partnerships. *Statistics in Medicine, 30*(7), 695–708. https://doi.org/10.1002/sim.4144.Strengthening

Machekano, R., et al. (2015). Meeting report: Workshop report: building biostatistics capacity in Sub-Saharan Africa-taking action. *Pan African Medical Journal, 21*(67). https://doi.org/10.11604/pamj.2015.21.167.5827

Machekano, R., et al. (2016). The Africa Center for Biostatistical Excellence: A proposal for enhancing biostatistics capacity for Sub-Saharan Africa. *Statistics in Medicine, 34*(27), 3481–3489. https://doi.org/10.1002/sim.6572.The

Ngianga-Bakwin, K., & Lawrence, K. (Eds.). (2019). *Statistical modelling of complex correlated and clustered data household surveys in Africa*. NOVA.

Thomson, D. R., et al. (2016). Applied statistical training to strengthen analysis and health research capacity in Rwanda. *Health Research Policy and Systems*, 1–7. https://doi.org/10.1186/s12961-016-0144-x

Part I
Data Harmonization and Analysis

Diagonal Reference Modelling of the Effects of Educational Differences Between Couples on Women's Health-Care Utilization in Eritrea

Gebrenegus Ghilagaber

Abstract We examine effects of differences in education between couples on women's propensity to utilize health care (specifically deliver at health facilities instead of at home). We contrast results from conventional logistic regression with those from diagonal reference models (DRM). Data used for illustration come from the 2002 Eritrean Demographic and Health Survey (DHS) and consist of 4255 women who have borne at least one child by the survey time (with a total of 6366 children). Standard logistic regression models indicate strong effects of educational differences on women's decision to deliver at health facilities. On the contrary, results from Diagonal Reference Modelling which accounts for origin (woman's education), destination (partner's education), and mobility (differences between couples' educational levels) show that there is no mobility effect. If any, DRM reveals that woman's own education is more important than her partner's education in such decisions. That the mobility effect disappeared in the DRM is in accordance with previous studies that used DRM and where mobility had no effect on the outcome variable in diverse fields. But, our recommendation is not to encourage users to use the results from DRM. Rather, we recommend to examine the DRM model more closely in the light of recent studies suggesting that an artifact of the model itself may lie behind the lack of mobility effects.

Keywords Social mobility · Diagonal reference models (DRM) · Health care utilization · Eritrea · Mother's education · Partner's education · Couple's education · Health decision · Hospital delivery · Birth outcome · Logistic regression · Binary outcome · Ideal number of children · Intended number of children · Family size · Maternal health · Prenatal care · Model artifact · Demographic and Health Surveys (DHS) · The DHS program · Developing countries · Family planning · Baseline levels

G. Ghilagaber (✉)
Stockholm University, Department of Statistics, Stockholm, Sweden
e-mail: Gebre@stat.su.se

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_2

1 Introduction

Institutional delivery and prenatal care have been some of the main recommendations of the World Health Organization (WHO) in order to improve the outcome of pregnancy in developing countries, see, for instance, World-Health-Organization and Others (2016). Since there are no formal randomized experiments on health care utilization, it is difficult to evaluate its benefits without examining correlates to health care utilization—especially in countries where health care centers may not be uniformly distributed across regions or rural and urban areas, see Ghilagaber (2014) for a method to correct for such selection bias.

To address this issue, investigators have often controlled for some of women's characteristics such as residence and education (alone or together with her partner's education and/or occupation) in models relating health inputs to outcome such as maternal health. However, the role of educational difference between couples has been ignored—especially in investigations based on data from Demographic and Health Surveys (DHS).

In this chapter, we attempt to fill this gap in knowledge and adapt models developed in the social mobility literature to examine effects of differences between couples' educational levels on women's propensity to utilize facilities (especially deliver a child at health facilities). These models, developed by Sobel (1981, 1985) and known as Diagonal Reference Models (DRM), account for woman's education (origin), her partner's education (destination) as well as differences between their educational levels (mobility). The model is applied on data from 4255 women with at least one child in the 2002 Eritrean Demographic and Health Survey (DHS) which resulted in 6366 children in total.

The main research question we intend to address is if educational mobility (being married to a partner with a different educational level) itself affects a woman's propensity to deliver a child at health facilities aside from the effects of her own educational level.

Results from conventional logistic regression reveal strong effects of mobility (educational differences) on women's decision to deliver at health facilities. However, such strong effects disappear when the same data is analyzed using the Diagonal Reference Models. Our results are in accordance with many other findings based on DRM in various fields and different outcome variables. Some recent works are reported in van der Waal et al. (2017), Boylan et al. (2014), Chaparro and Koupil (2014), Aitsi-Selmi et al. (2013), Kuntz and Lampert (2013), Krzyżanowska and Mascie-Taylor (2011), and Heraclides and Brunner (2010). But, we are not yet in a position to recommend the DRM model because there are other ongoing studies suggesting that the lack of mobility effects in the DRM can be an artifact of the model itself.

We introduce our illustrative data set in Sect. 2. In Sect. 3, we describe the Diagonal Reference Model. This model is then applied on our data set in Sect. 4 and the results are compared with those from a standard logistic regression model for the propensity to deliver in health facilities. We summarize the findings of our chapter by way of concluding remarks and recommendations in Sect. 5.

2 Data: Hospital Delivery Among Eritrean Women

Data used for illustration in this chapter come from the 2002 Eritrean Demographic and Health Survey, see National-Statistics-Office-Eritrea and Macro-International-Inc. (2003). Usable records for the purpose of this chapter consist of 4255 women who have borne at least one child by the survey time (March–July 2002) and with valid values for the their own and partner's educational levels and other background variables. The total number of births by these women was 6366 children.

Frequency distribution of the data set across couples' educational levels is shown in Table 1 (left panel) while the corresponding percent distribution is shown on the right panel. Thus, 2278 women (53.54%) had no education and were married to a partner with no education while 448 women (10.53%) reported they have primary-level education and were married to a partner with the same level of education. 401 of the 4255 women (9.42%) reported to have secondary or higher level education but only 32 (0.75% of the entire sample) were married to a partner with no education.

Table 2 displays frequency distribution of the women who reported to have delivered their first child in hospital. Thus, only 1034 of the 4255 women (24.30%) delivered their first child at health facilities while the rest 75.70% reported that they have delivered their first child at home (with the help of traditional midwives). The right panel of Table 2 shows corresponding percent distributions of these 1034 hospital deliveries.

Table 3 shows percent of hospital deliveries among all women and is obtained by dividing the frequencies in Table 2 by the corresponding entry in Table 1. The educational gradient in the propensity to deliver at hospital is clear in Table 3. Thus, while couples with no education constitute more than half of the sample (53.54%), only 8.43% of women in this group have delivered in hospital. Couples with the

Table 1 Frequency and percent distribution of the sample across couples' education

Frequencies	Partner's educ				Percentages	Partner's educ			
Own educ	No	Prim.	Sec.	Total	Own educ	No	Prim.	Sec.	Total
No Educ	2278	591	81	2950	None	53.54	13.89	1.90	69.63
Primary	209	448	247	904	Primary	4.91	10.53	5.80	21.25
Second+	32	66	303	401	Second+	0.75	1.55	7.12	9.42
Total	2519	1105	631	4255	Total	59.20	25.97	14.83	100

Table 2 Frequency and percent of hospital delivery across couples' education

Frequencies	Partner's educ				Percentages	Partner's educ			
Own educ	No	Prim.	Sec.	Total	Own educ	No	Prim.	Sec.	Total
No Educ	192	95	38	325	None	18.57	9.19	3.68	31.43
Primary	61	155	151	367	Primary	5.90	14.99	14.60	35.49
Second+	25	51	266	342	Second+	2.42	4.93	25.73	33.08
Total	278	301	455	1034	Total	26.89	29.11	44.00	100

Table 3 Percentage of hospital deliveries in the sample

	Partner's educ			
Woman's educ	No Educ	Primary	Secon+	Total
No Educ	8.43	16.07	46.91	11.02
Primary	29.19	34.60	61.13	40.60
Second+	78.13	77.27	87.79	85.29
Total	11.04	27.24	72.11	24.30

highest educational level (secondary or higher), on the other hand, constitute only 7.12% of the sample but 87.79% of women in this group have reported they delivered their first child at health facilities.

Overall we observe a strong association between couples' educational levels and the decision to deliver in health facilities (hospital) but also that women's own education is more important than partner's education in such decisions. As reported in Ghilagaber (2018) part of the explanation may be that highly educated couples live in urban areas where health facilities are easily accessible and we will control for residence and other background variables when analyzing the data in Sect. 4 after introducing the Diagonal Reference Model below.

3 Diagonal Reference Models

The diagonal reference model was developed by Sobel (1981, 1985) to build on the rectangular model of Hope (1971) to examine the effect of social mobility (changes in social class between generations) on demographic outcome. It was later applied to other areas like political efficacy in Clifford and Heath (1993), attitude towards immigrants in Paskov et al. (2019), health and well-being in Chan (2018) and Präg and Richards (2019). It was also adapted to problems outside social class such as educational differences in, among others, Eeckhaut et al. (2013).

The diagonal reference model treats origin, destination, and indicators for upward and downward mobility differently. In the original formulation in Sobel (1981, 1985), origin refers to parent's social class, destination to own social class, and mobility as belonging to a social class that is different from parent's. In this chapter, however, origin refers to a woman's own educational level, destination refers to her partner's educational level, and mobility refers to having a partner with different educational level hers.

The model is designed for contingency tables classified by factors with the same levels (same number of rows and columns). The main diagonal then represents the "immobile" individuals (those who have the same educational level as their partners) and are assumed to set the norm of behavior. The mobile individuals occupy the off-diagonal cells and are either upward mobile (their own educational level is lower than their partner's) or downward mobile (their own educational level is higher than their partner's).

Turner and Firth (2022) model the cell means as a function of the diagonal effects, i.e., the mean responses of the "diagonal" cells in which the levels of the row and column factors are the same. In our present case, we have a three-way square matrix defined by the three educational levels described in the previous section (No Educ, Primary, Secondary, or higher). Following Turner and Firth (2022), if the mean response in cell (i, j) is denoted by μ_{ij}, then the diagonal reference model expresses it as

$$\mu_{ij} = \omega\mu_{ii} + (1 - \omega)\,\mu_{jj},$$

where ω $(0 < \omega < 1)$ is a weight associated with the origin (woman's education) and reflects the degree of importance of her own education in couples with different educational levels. Women in cell (i, i) and their partners in cell (j, j) represent "pure" i and j effects, respectively, whereas individuals in cells (i, j) have partners with lower or higher educational levels than their own and, hence, represent some intermediate category.

According to Turner and Firth (2022), a diagonal reference term comprises an additive component for each factor. The component for factor f is given by

$$\omega_f = \frac{\exp(\delta_r)}{\sum_r \exp(\delta_r)},$$

where the sum is over the levels of the factor and δ_r is a parameter to be estimated.

Thus, in a diagonal reference model for a contingency table classified by the row factor i (origin = woman's own education) and the column factor j (destination = her partner's educations), the mean response in cell (i, j) is given by

$$\mu_{ij} = \omega\gamma_i + (1 - \omega)\gamma_j = \left(\frac{\exp(\delta_1)}{\exp(\delta_1) + \exp(\delta_2)}\right)\gamma_i + \left(\frac{\exp(\delta_2)}{\exp(\delta_1) + \exp(\delta_2)}\right)\gamma_j,$$

where γ_i and γ_j are mean responses of origin i and destination j (ith education level of woman and jth education level of her partner) and δ_1 and δ_2 are parameters to be estimated.

In the presence of one or more explanatory variables, as is the case in our illustrative example where we control for four background variables (birth cohort, region, residence, and ethnicity), the above model may be extended as follows, see van der Slik et al. (2002) and Turner and Firth (2022):

$$\mu_{ijk} = \beta_1 x_{1k} + \beta_2 x_{2k} + \beta_3 x_{3k} + \beta_4 x_{4k} + \left(\frac{\exp(\delta_1)}{\exp(\delta_1) + \exp(\delta_2)}\right)\gamma_i + \left(\frac{\exp(\delta_2)}{\exp(\delta_1) + \exp(\delta_2)}\right)\gamma_j.$$

Thus, the problem reduces to estimating the parameters δ_1 and δ_2 (and from them the weight ω from the relation shown in the equation for ω above) and the covariate effects β if the model includes explanatory variables. This is achieved using the Dref

option in the R-package for general nonlinear models, gnm, developed by Turner and Firth (2022).

According to Turner and Firth (2022) the diagonal effects represent contrasts with the off-diagonal cells and, hence, do not need to be constrained. Further, the coefficients of the covariates are not aliased with the parameters of the diagonal reference term implying it suffices with the usual constraints (using one of the levels as baseline reference category). The only unidentified parameters in the DRM model are the weight parameters, ω_i.

In the next section we fit the above model to our data set described in Sect. 2 and compare the results with those obtained from a standard logistic regression model for the decision to deliver a child at a health facility.

4 Application: Educational-Mobility Effects on Hospital Delivery

4.1 Measures

Our response variable is the binary outcome on whether a woman delivers her first child at health facilities (hospital or clinics) rather than at home. We have access to women with multiple children but we concentrate on the first birth in this chapter. The main rational behind this choice is to avoid correlation among children from the same mother and, hence, underestimation of standard errors of covariates which, in turn, would lead to spurious significance.

Our main explanatory variable is educational mobility which, in our case, refers to women whose educational level is different from their partners'. In addition to educational mobility, women's own education (origin) and their partners' education (destination) are part of the model. We also controlled for 5 background variables—birth cohort, region, residence (urban or rural), religion, and ethnicity. But, in our results section below we will report only those of primary interest (origin, destination, and mobility).

4.2 Results

4.2.1 Results from Conventional Logistic Regression

The presentation of our results in this section follows that of van der Waal et al. (2017). Since we have a 3 by 3 contingency table, we have 9 educational groups (pairs of own-partner education). In the conventional logistic regression, one of these groups (couples without education) is used as a baseline level. This produces 8 dummy variables whose estimated parameters are shown in the upper panel of Table 4 together with their corresponding standard errors, odds ratios, and 95%

Table 4 Estimated mobility effects from conventional logistic regression and two DR models

Model	Estimate	Stand. error	OR	Lower 95%	Upper 95%
Conventional logistic regression					
No Educ married to men with Prim. Educ	0.359	0.129	1.43	1.11	1.84
No Educ married to men with Sec. Educ	1.159	0.231	3.19	2.03	5.01
Prim. Educ married to men with No Educ	0.856	0.171	2.35	1.68	3.29
Prim. Educ married to men with Prim. Educ	0.976	0.133	2.65	2.05	3.44
Prim. educ married to men with Sec. Educ	1.454	0.157	4.28	3.14	5.82
Sec. Educ married to men with No Educ	1.911	0.394	6.76	3.12	14.64
Sec. Educ married to men with Prim. Educ	1.965	0.278	7.14	4.14	12.30
Sec. Educ married to men with Sec. Educ	2.761	0.205	15.82	10.59	23.64
DRM (ref: Downward mobile)					
"Immobile" couples with no education	−0.595	0.286	0.55	0.31	0.97
"Immobile" couples with primary education	−0.095	0.278	0.91	0.53	1.57
"Immobile" couples with secon+ education	−0.089	0.322	0.91	0.49	1.72
Upward (own educ lower than partner's)	−0.454	0.247	0.64	0.39	1.03
DRM (ref: Upward mobile)					
"Immobile" couples with no education	−0.141	0.241	0.87	0.54	1.39
"Immobile" couples with primary education	0.359	0.268	1.43	0.85	2.42
"Immobile" couples with secon+ education	0.365	0.360	1.44	0.71	2.92
Downward (own educ higher than partner's)	0.454	0.247	1.57	0.97	2.55

confidence intervals. We see that all 8 educational combinations have higher odds of delivering a child in hospital compared to the baseline group of couples with no education. The confidence intervals for the odds ratios show that the differences are significant at 5% significance level (in fact the p-values are all less than 0.01 except for the first group whose p-value is between 0.01 and 0.05).

The odds ratios in the upper panel of Table 4 are reproduced in the 3 by 3 array on the left panel of Table 5. From these we produced profiles of odds ratios across woman's education (middle panel in Table 5) and partner's education (right panel

Table 5 OR (left) and OR-profiles across own (mid) and partner's educ (right)

Educ	Partner's			Educ	Partner's			Educ	Partner's		
Own	No	Prim.	Sec.	Own	No	Prim.	Sec.	Own	No	Prim.	Sec.
No	1	1.43	3.19	No	1	1	1	No	1	1.43	3.19
Prim.	2.35	2.65	4.28	Prim.	2.35	1.85	1.34	Prim.	1	1.13	1.82
Sec.	6.76	7.14	15.82	Sec.	6.76	4.99	4.96	Sec.	1	1.06	2.34

in Table 5). From Table 5 (left panel), we note that the odds of hospital delivery are about 16 times among couples with secondary or higher education compared to couples with no education. This combination of highest education is also confirmed in the last row of the middle panel where women with secondary or higher education have the highest odds of hospital delivery. The last column of the right panel where women whose partners have secondary or higher education have the highest odds also lends support to the confirmation.

A result worth noting Table 5 is that among women with the highest education (middle panel) it is those women whose partners have no education who have the highest odds (though the differences are not large ranging between 4.96 to 6.76). Similarly, we note that among women whose partners have the highest education (right panel), it is those with no education who have the highest odds (3.19 compared to 1.82 and 2.34).

Another interesting result in Table 5 is that women's own education seems to have stronger effect on the decision to deliver at hospital compared to their partners' education. We will assess this formally in the results for DRM.

4.2.2 Results from Diagonal Reference Models (DRM)

Results from two variants of the diagonal reference model are shown in the middle and lower panels of Table 4. In the middle panel (where downward mobile women are treated as baseline) we note that the mean effects of the diagonal elements are estimated as $\mu_{11} = -0.595$, $\mu_{22} = -0.095$, and $\mu_{33} = -0.089$, while the estimate associated with upward mobility is –0.454.

The results in the lower panel of Table 4 come from a DRM where upward mobile women are treated as baseline. The results are now in the opposite directions except for μ_{11}. In fact, the estimate associated with downward mobility (0.454) is just the negative of that for upward mobility in the middle panel.

But, the most striking result in the two DRM models is that only couples with no education have significantly lower odds of hospital delivery than the baseline couple (where women's education is higher than their partners'). None of the other effects is significant as indicated by their corresponding 95% confidence intervals for the odds ratios.

The weights associated with the origin (woman's own education) and destination (her partner's education) are displayed in Table 6 for our response variable (hospital

Table 6 Estimated origin and destination weights in Diagonal Reference Models for three health related outcomes

Response variable	Origin (Woman's educ)	Destination (Partner's educ)
Delivery at health facilities	0.54	0.46
Intended number of children	0.30	0.70
Ideal number of children	0.58	0.42

delivery) and two other health related outcomes in the same survey (intended number of children and ideal number children, dichotomized into small and large).

The weights, ω, reflect the extent mobile women are influenced by origin effects (their own educational level) relative to destination effects (their partners' educational level). Thus, the results for hospital delivery show that woman's education is more important ($\omega = 0.54$) than her partner's education ($1 - \omega = 0.46$) though the difference is not large. This can be compared with the estimate of the weight for another outcome variable (intended number of children). For this outcome, partner's education ($1 - \omega = 0.70$) weighs more than twice her own education ($\omega = 0.30$).

Thus, our results from the DRM indicate that for women whose education is different from their partners', the decision to deliver at health facilities is influenced almost equally by their own and their partners education. Further, after accounting for women's own and their partners' education, there is no effect of educational differences between couples on the decision to deliver at health facilities.

5 Summary and Concluding Remarks

In this chapter, we explored the relationship between educational mobility and the propensity to deliver at health facilities in Eritrea based on data from its 2002 Demographic and Health Survey.

Previous studies have considered woman's educational level alone or together with partner's education but ignored effect of educational differences. Thus, the scientific question we intended to address was if differences in educational levels between couples affect woman's decision to utilize health care (specifically deliver a child at health institution).

We used the diagonal reference model which has been recommended in the literature because, it is argued, can accurately capture the effect of mobility and isolate it from those of origin and destination. Propensity to deliver a child at health centers was modelled as a function of women's own and their partners' education as well as educational differences between the couples. We also controlled for some background variables (birth cohort, region, residence, religion, and ethnicity).

Results from conventional logistic regression models showed that there is strong association between educational differences among couples and women's propensity to deliver at hospital. However, these strong associations disappeared

when the data was analyzed using Diagonal reference Models (DRM). These results are consistent with findings on the application of DRM in various fields (that mobility has no effect on a range of important outcomes).

Recent works by Fosse and Pfeffer (2019) caution researchers who use DRM that the resulting estimated mobility effects can, in part, be an artifact of the model. We, therefore, suggest that future studies focus on a closer look at the DRM model and its properties before recommending its universal use.

Acknowledgments The data sets analyzed in this chapter were provided by the Eritrean National Statistics Office (NSO) and the Demographic and Health Surveys (DHS) program. We are grateful to the DHS program and to the director of NSO, Mr. Ainom Berhane, and its senior statisticians, Mr. Hagos Ahmed and Mr. Samuel Tesfamariam. The views expressed in the chapter are solely of the author and do not express the views or opinions of the data sources or their employees.

Appendix: R-Codes Used for Computing the Results in Some of the Tables in This Chapter

```
-------------------------------------------------------------------------

  Installing necessary packages, reading the source file in excel
format, and defining the origin and destination.

-------------------------------------------------------------------------

  install.packages("xlsx")
  library(readxl)
  install.packages("gnm")
  require(gnm)
  Eri2002 <- readexcel("C:/DRM-Revised-Educ-Occup-Hosp-Children-Ideal.
  xlsx")
  View(Eri2002)
  Origin<-factor(Eri2002$WomEdu)
  Destination<-factor(Eri2002$HusbEdu)
  The source file has 6366 rows (children) from 4255 mothers and 24
  columns(woman´s education, husband´s education, categorical
  mobility indicators, and many other background variables). But, not
  all columns are used in the analyses.

-------------------------------------------------------------------------

  Fitting conventional logistic regression models on the binary
outcome (hospital delivery) using some background factors as
covariates.

-------------------------------------------------------------------------

HospConv1 <- gnm(Delivery ~ -1 + factor(Cohort) + factor(Region) +
factor(Resid) + factor(Religion) + factor(Ethn) + factor(EduMob),
family = binomial, data = Eri2002)
  summary(HospConv1)
  HospConv2 <- gnm(Delivery ~ -1 + factor(Cohort) + factor(Region) +
```

```
factor(Resid) + factor(Religion) + factor(Ethn) + factor(ImmobPrim) +
factor(ImmobSecon)+ factor(Upwards) + factor(Downwards), family =
binomial, data = Eri2002)summary(HospConv2)

------------------------------------------------------------------------

------------------------------------------------------------------------

  Fitting Diagonal Reference Model for Hospital Delivery with
downwards mobile women as baseline (reference) in mobility.

------------------------------------------------------------------------

  HospUp <- gnm(Delivery ~ -1 + factor(Cohort) + factor(Region) +
factor(Resid) + factor(Religion) + factor(Ethn) + factor(Immobile) +
factor(Upwards) + Dref(Origin,Destination), family = binomial, data =
Eri2002)
  summary(HospUp)
  DrefWeights(HospUp)

------------------------------------------------------------------------

  Fitting Diagonal Reference Model for Hospital Delivery with upwards
mobile women as baseline (reference) in mobility.

------------------------------------------------------------------------

  HospDown <- gnm(Delivery ~ -1 + factor(Cohort) + factor(Region)
+ factor(Resid) + factor(Religion) + factor(Ethn) + factor(Immobile) +
factor(Downwards) + Dref(Origin, Destination), family = binomial, data
= Eri2002)
  summary(HospDown)
  DrefWeights(HospDown)

------------------------------------------------------------------------
```

References

Aitsi-Selmi, A., Batty, G., Barbieri, M., Silva, A., Cardoso, V., Goldani, M., Marmot, M., & Bettiol, H. (2013). Childhood socioeconomic position, adult socioeconomic position and social mobility in relation to markers of adiposity in early adulthood: Evidence of differential effects by gender in the 1978/79 ribeirao preto cohort study. *International Journal of Obesity, 37*(3), 439–447.

Boylan, S., Gill, T., Hare-Bruun, H., Andersen, L., & Heitmann, B. (2014). Associations between adolescent and adult socioeconomic status and risk of obesity and overweight in Danish adults. *Obesity Research and Clinical Practice, 8*(2), e163–e171.

Chan, T. (2018). Social mobility and the well-being of individuals. *The British Journal of Sociology, 69*(1), 183–206.

Chaparro, M., & Koupil, I. (2014). The impact of parental educational trajectories on their adult offspring's overweight/obesity status: A study of three generations of Swedish men and women. *Social Science and Medicine, 120*, 199–207.

Clifford, P., & Heath, A. (1993). The political consequences of social mobility. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 156*(1), 51–61.

Eeckhaut, M., Van de Putte, B., Gerris, J., & Vermulst, A. (2013). Analysing the effect of educational differences between partners: A methodological/theoretical comparison. *European Sociological Review, 29*(1), 60–73.

Fosse, E., & Pfeffer, F. (2019). Beyond the diagonal reference model: Critiques and new directions. In *2019 Annual Meeting of the Population Association of America (PAA)*, Austin, TX, 10–13 April 2019. http://paa2019.populationassociation.org/abstracts/190797.

Ghilagaber, G. (2014). Disentangling selection and causality in assessing the effects of health inputs on child survival: Evidence from east Africa. In *Advanced Techniques for Modelling Maternal and Child Health in Africa*. Springer Series on Demographic Methods and Population Analysis (Vol. 34, pp. 11–28).

Ghilagaber, G. (2018). Statistical analysis of educational and urban-rural differences in marriage-timing: Evidence from the 2010 Eritrean population and health survey for women and men. In *Proceedings of the International Conference on Eritrean Studies*, 20–22 July 2016 (pp. 1097–1124).

Heraclides, A., & Brunner, E. (2010). Social mobility and social accumulation across the life course in relation to adult overweight and obesity: The Whitehall II study. *Journal of Epidemiology and Community Health, 64*(8), 714–719.

Hope, K. (1971). Social mobility and fertility. *American Sociological Review*, 1019–1032.

Krzyżanowska, M., & Mascie-Taylor, C. (2011). Intra-and intergenerational social mobility in relation to height, weight and body mass index in a British national cohort. *Journal of Biosocial Science, 43*(5), 611–618.

Kuntz, B., & Lampert, T. (2013). Intergenerational educational mobility and obesity in adolescence: Findings from the cross-sectional German KIGGS study. *Journal of Public Health, 21*(1), 49–56.

National-Statistics-Office-Eritrea and Macro-International-Inc. (2003). *Eritrea demographic and health survey 2002*. National Statistics and Evaluation Office and ORC Macro.

Paskov, M., Präg, P., & Richards, L. (2019). Does downward social mobility make people more hostile towards immigrants?.

Präg, P., & Richards, L. (2019). Intergenerational social mobility and allostatic load in great Britain. *J Epidemiol Community Health, 73*(2), 100–105.

Sobel, M. (1981). Diagonal mobility models: A substantively motivated class of designs for the analysis of mobility effects. *American Sociological Review*, 893–906.

Sobel, M. (1985). Social mobility and fertility revisited: Some new models for the analysis of the mobility effects hypothesis. *American Sociological Review*, 699–712.

Turner, H., & Firth, D. (2022). Package 'gnm' (Generalized Nonlinear Models) - Version 1.1-2 (29 April 2022). The Comprehensive R Archive Network Project. https://cran.r-project.org/.

van der Slik, F., De Graaf, N., & Gerris, J. (2002). Conformity to parental rules: Asymmetric influences of father's and mother's levels of education. *European Sociological Review, 18*(4), 489–502.

van der Waal, J., Daenekindt, S., & de Koster, W. (2017). Statistical challenges in modelling the health consequences of social mobility: The need for diagonal reference models. *International Journal of Public Health, 62*(9), 1029–1037.

World-Health-Organization and Others. (2016). *WHO recommendations on antenatal care for a positive pregnancy experience*. World Health Organization.

Sequential Probit Modeling of Regional Differences in the Effects of Education on Parity Progression Ratios in Ethiopia

Gebrenegus Ghilagaber and Paraskevi Peristera

Abstract A sequential probit model is applied to analyze differentials in the effects of women's educational level on parity progression ratios in Ethiopia. Since parity progression requires successful completion of the prior parity for progression into the next higher parity, we argue that a sequential decision model captures the decision process more accurately. Further, since reasons to have a first child may differ from those to have, say, a second or third child, we allow the effects of covariates on the progression propensities to vary between parities in the same model. Data used for illustration come from the Ethiopian Mini Demographic and Health Survey of 2019 in which 8885 women from 11 regions were interviewed. Results show that the sequential model provides more insight than conventional models when exploring the association between education and parity progression in particular and fertility decision process in general. We also found both similarities and differences in the effects of education on parity progression among the regions. We included a household random-effect term to account for women's clustering within households. The random effect was significant in a model for the entire country but disappeared when region was included as a covariate in the model.

Keywords Sequential models · Probit model · Sequential decision process · Fertility · Parity · Parity progression · Childlessness · Children ever born · Propensity · Family planning · Family size · Education · Educational gradients of parity progression and childlessness · Baseline levels · Demographic and Health Surveys (DHS) · Regional differences · Household random effects · Unobserved heterogeneity · Stratified modelling · Clustered data · Multilevel modelling · Ethiopia · Tigray region · Afar region · Amhara region · Oromia

G. Ghilagaber (✉)
Stockholm University, Department of Statistics, Stockholm, Sweden
e-mail: Gebre@stat.su.se

P. Peristera
Stockholm University, Stress Research Institute, Department of Psychology, Stockholm, Sweden
e-mail: Paraskevi.Peristera@su.se

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_3

region · Somali region · Benishangul Gumuz region · Southern nations · Nationalities and peoples region · Gambela region · Harari region · Addis Abeba city administration · Dire Dawa city administration

1 Introduction

Over the past four decades, the Demographic and Health Survey (DHS) program has played a vital role in conducting nationally representative household surveys in developing countries. Among the information gathered by DHS is the information on current and total fertility such as age at first marriage and first birth, birth intervals by background variables, children ever born, and fertility preferences as indicated by an ideal number of children. Apart from their strong association to maternal and child health (for instance, shorter birth intervals increase childhood mortality; age at childbirth can affect maternal health and birth outcome), the above indicators are important for monitoring population growth and developing family planning programs. A notable example is the 1970s Chinese policy for limiting fertility. The policy was based on the slogan "wan, xi, shao" ("later, longer, fewer") and strongly promoted later marriage, longer birth intervals, and fewer children in total.

The rich data in Demographic and Health Surveys (DHS) enable investigators to make in-depth analyses that guide policy intervention. Such analyses, in turn, require advanced statistical techniques in order to get maximum use of the available data.

The effect of timing of age at first marriage and first birth on fertility has been documented in, among others, Marini (1981). Recently, Arroyo et al. (2017) and Eickmeyer et al. (2017) studied changes in median ages at first marriage and first birth over the period 1980–2017. Gurmu and Etana (2014) analyze the roles of social and demographic factors on age at first marriage in Ethiopia.

A birth-interval approach to the study of fertility was studied in, for instance, Ghilagaber et al. (2005) where models were proposed for the quantum of fertility (the proportion of women who move to the next higher parity) and the tempo of fertility (the time it takes to make the progression for those women who continue reproduction). Accelerated failure-time models for the tempo of fertility and dynamic survival models for the quantum of fertility are also presented elsewhere in this book and illustrated with data on age at first marriage in a Bayesian framework in Liang and Ghilagaber (2022); Munezero and Ghilagaber (2022) and with data on age at first birth in Ghilagaber et al. (2022).

The above sample of studies address the "later" and "longer" components of fertility-related policy (in the context of the Chinese slogan). In the present chapter, we propose and apply a method for assessing the last component ("fewer") of the slogan. The information we extract is the total number of children reported by women interviewed in the Demographic and Health Surveys (DHS). But, rather than applying methods for count data on the number of children, we propose sequential procedure for the conditional propensity to progress to the next higher parity for

women who have completed a given parity. Our approach is closer to the quantum approach to fertility because our interest is in the proportion of women who move to the next higher parity.

We illustrate our approach using data on parity progression among women in Ethiopia based on data from the 2019 Mini DHS in the country, see Ethiopian-Public-Health-Institute and ICF (2019). The information on which our response variable is based is the total number of children borne by the respondent by the survey time. This is sequentially ordered with values between 0 and 15.

It may be tempting to analyze such data by dichotomizing the number of children into two groups (small and large) with some threshold for what is a large number of children and applying conventional logistic regression models. We argue, in accordance with Amemiya (1978), Maddala (1986), Mare (1980), Nagakura (2004), and Waelbroeck (2005), that such procedure is subjective, and different conclusions can be reached for different choices of thresholds. Instead, we propose sequential probit modeling that is appropriate for outcome variables that are ordered sequentially, which is the case in our data set. For more applications of sequential models in different areas, see, for instance, Alpu and Fidan (2004); Amemiya (1985); Brien and Lillard (1994); Davidson and MacKinnon (2004); Liao (1994); Munkin (2011); Steele and Durrant (2011). Ghilagaber and Peristera (2014) use multilevel sequential probit to model neighborhood effects on educational progress among children to Polish and Turkish immigrants in Sweden.

We present the probit and sequential probit models in the next section. In Sect. 3, we present our data set, apply the models on the data set, and present the results. We summarize our findings together with some concluding remarks in Sect. 4.

2 Probit and Sequential Probit Models

2.1 Probit Model

Following Albert (2009), let us present a woman's decision to progress to the next higher parity by a binary indicator variable Y_i where $Y_i = 1$ if the woman decides to progress to the next higher parity and $Y_i = 0$ if she does not.

Suppose there exists a continuous measurement Z_i of decision such that Z_i is positive if woman i decides to progress to the next higher parity and Z_i is negative if woman i does not make the progress. Moreover, the decision measurement is related to the k covariates $x_{i1}...x_{ik}$ (such as k dummy variables indicating different levels of education) by the normal regression model:

$$Z_i = x_{i1}\beta_1 + ... + x_{ik}\beta_k + \epsilon_i,$$

where ϵ_i, ..., ϵ_n is a vector of error terms from the standard normal distribution.

The probit regression model (which is analogous to the logistic regression model) expresses the probability $p_i = P(Y_i = 1)$ as

$$p_i = P(Y_i = 1) = P(Z_i > 0) = \Phi(x_{i1}\beta_1 + ... + x_{ik}\beta_k),$$

where $(\beta_1, ..., \beta_k)$ is a vector of unknown parameters and Φ is the cumulative distribution function of a standard normal distribution.

2.2 Sequential Probit Model

Suppose one observes N independent women and W_i is the outcome variable with J possible ordered values $\{j = 1, \ldots, J\}$. Let $x_i = (x_{i1}, \ldots x_{ik})$ denote a set of k covariates associated with response W_i.

In the sequential model, the variable W_i can take the value j only after the levels $1, \ldots, j-1$ are reached. So, in order to get the outcome j, one must first have experienced levels $1, 2, \ldots, j-1$. The conditional probability of reaching level j $(1 \leq j \leq J-1)$ is given by

$$\Pr(W_i = j | W_i \geq j, \gamma, \delta) = \Phi\left(\gamma_j - x_i^{'}\delta\right), \tag{1}$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution, δ is the regression parameter vector, $\gamma = (\gamma_1, \ldots, \gamma_{J-1})$ are threshold parameters, and $x_i^{'}\delta$ represents the effect of covariates. The unconditional probabilities are defined as follows:

- The probability of reaching level j is given by

$$\Pr(W_i = j | \gamma, \delta) = F\left(\gamma_j - x_i^{'}\delta\right) \prod_{r=1}^{j-1} \left\{1 - \Phi\left(\gamma_r - x_i^{'}\delta\right)\right\}, \; j \leq J-1. \tag{2}$$

- The probability of reaching the highest level J is given by

$$\Pr(W_i = J | \gamma, \delta) = \prod_{r=1}^{J-1} \left\{1 - \Phi\left(\gamma_r - x_i^{'}\delta\right)\right\}. \tag{3}$$

The sequential model is often formulated in terms of latent variables. Let us define $\{u_{ij}\}$ the latent variables corresponding to the i-th observation,

$$\{u_{ij}\} = x_i^{'}\delta + e_{ij}, \tag{4}$$

where e_{ij} are independently distributed from Φ. The observed data are then obtained as

$$W_i = \begin{cases} 1 & if\ u_{i1} \leq \gamma_1 \\ 2 & if\ u_{i1} > \gamma_1, u_{i2} \leq \gamma_2 \\ \vdots & \vdots \\ j\ if\ u_{i1} > \gamma_1, \ldots, u_{ij-1} > \gamma_{j-1}, u_{ij} \leq \gamma_j \end{cases}. \tag{5}$$

In the above model, the latent variable u_{ij} represents a woman's propensity to progress to parity $j + 1$, given that she already has j children. This implies that sequential models can be estimated by conditioning on the appropriate sub-samples in the data.

2.3 Parameter Estimation in the Sequential Probit Model

In classical framework, the parameters in the sequential probit model can be estimated using the maximum likelihood approach by maximizing the likelihood function,

$$\begin{aligned} L(\gamma, \delta) = & \prod_{i:y_i<J} \left[\Phi\left(\gamma_{y_i} - x_i^{'}\delta\right) \prod_{r=1}^{y_i-1} \left\{1 - \Phi\left(\gamma_r - x_i^{'}\delta\right)\right\} \right] \\ & * \prod_{i:y_i=J} \left[\prod_{r=1}^{J-1} \left\{1 - \Phi\left(\gamma_r - x_i^{'}\delta\right)\right\} \right]. \end{aligned} \tag{6}$$

Fahrmeir and Tutz (2001) have shown that sequential ordered models are special cases of multivariate generalized linear models and that maximum likelihood estimates can be obtained using an iterative re-weighted least-squares algorithm.

For more details on estimation in sequential ordered models, we refer the reader to chapter 3 of Fahrmeir and Tutz (2001), where several sequential models with different choices of distribution functions (the function F in Eq. 2 above) are presented. We also refer to chapter 5 of Lillard and Panis (2003) on which the program we use for computation in this chapter (aML) is based. Markov chain Monte Carlo (MCMC) algorithms have also been developed for fitting such models. See, for instance, Albert and Chib (2001) and Waelbroeck (2005).

3 Application to Parity Progression Among Ethiopian Women

3.1 Data Set and Measures

The data set for our illustration comes from the Ethiopia Mini Demographic and Health Survey of 2019, see Ethiopian-Public-Health-Institute and ICF (2019).

Our response variable is the propensity of progression to next higher parity among women who have completed the current parity. Thus, all women are considered in modeling progression from parity 0 (no child) to parenthood (parity 1). Those women who have become parents (have at least one child) are then considered in the next progression (from parity 1 to parity 2), while those who are still with no child are no longer considered in further analyses. Similarly, women with only one child are not considered in modeling progression from parity 2 to parity 3, and so on.

The program we use for computation is aML (Applied Multilevel) developed by Lillard and Panis (2003). It enables to estimate the parameters for multiple progressions from one model, and thus, one does not need to fit separate models for each progression.

Our major explanatory variable is mother's educational level (highest educational level attained by the survey time). Educational gradient of parity progression (including childlessness) is well documented in the literature. See, for instance, Wood et al. (2014) for study on 14 low-fertility countries. We also include region as explanatory variable while analyzing data for the entire country. Else, education alone was used as explanatory variable while analyzing data for each region separately. The goal of the illustration is to demonstrate the methodological contribution described in the chapter, and hence, we are less interested in the substantive demographic question on potential correlates of parity progression.

A distribution of a total number of children ever borne by the survey time, cross classified by educational level and region, is presented in Table 1. From the last column in the bottom panel of table, we see that overall 3039 of the 8885 women (34%) were childless by the survey time, 1146 women (13%) had one child, 1047 (12%) had two children, etc. We also note regional differences in that panel. While 444 of the 818 women (54%) from the capital city, Addis Ababa, and 343 of 812 women (42%) from the other city administration, Dire Dawa, were childless by the survey time, the corresponding figures for the Afar and Somali regions were 141 of 641 (22%) and 194 of 723 (27%), respectively.

The column totals at the bottom of Table 1 show that 733 of the 8885 women (8.25%) were from the Tigray region, 641 (7.21%) were from the Afar region, 948 (10.67%) were from the Amhara region, 818 (9.21%) were from the capital city, Addis Ababa, etc.

Further, adding the values in the last columns of each panel (educational level) shows that 3640 of the 8885 women (40.97%) had no education, 3345 (37.65%) had primary-level education, 1149 (12.93%) had secondary-level education, while the rest 751 (8.45%) had higher (above secondary) education. The last row of Table 1 shows the number of households in which the corresponding women were clustered in.

Educational differences in the number of children ever born are shown in Fig. 1 for each region as well as for the entire country (last plot in Fig. 1), while Fig. 2 shows regional differences in the number of children for each of the four educational

Table 1 Number of children ever borne by womens' educational level and region: Ethiopia Mini-DHS 2019

Educ.	Children	Tigray	Afar	Amhara	Oromia	Somali	Benish.	SNNPR	Gambela	Harari	Addis A.	Dire D.	Total
No education	0	30	55	30	39	110	27	33	14	42	41	43	464
	1	26	48	35	29	27	17	16	17	24	15	26	280
	2	28	56	69	28	35	29	36	16	22	17	26	362
	3	40	68	71	27	45	36	44	22	25	20	19	417
	4	42	53	77	50	41	43	58	30	25	4	20	443
	5	41	41	68	43	43	38	58	32	25	7	34	430
	6	32	35	45	58	55	40	61	35	24	4	18	407
	7	29	36	38	53	45	32	41	20	17	0	15	326
	8	14	21	23	29	18	31	38	13	13	2	16	218
	9	6	8	13	25	26	9	21	5	5	0	13	131
	10	4	7	5	15	13	8	12	0	9	0	9	82
	11	1	5	2	11	11	4	2	2	0	0	6	44
	12	1	1	5	5	2	4	2	1	1	0	1	23
	13	0	2	0	0	2	0	0	0	0	0	1	5
	14	1	0	0	1	1	0	2	0	0	0	1	6
	15	0	0	0	0	1	0	0	0	1	0	0	2
Primary	0	83	69	192	223	72	134	211	136	108	150	156	1534
	1	49	30	59	59	10	54	48	54	48	61	46	518
	2	32	33	36	73	12	38	46	41	42	33	35	421
	3	21	15	13	36	7	31	48	34	29	17	22	273

(continued)

Table 1 (continued)

Educ.	Children	Tigray	Afar	Amhara	Oromia	Somali	Benish.	SNNPR	Gambela	Harari	Addis A.	Dire D.	Total
	4	18	6	18	34	5	22	24	23	23	7	14	194
	5	16	5	11	29	6	15	23	18	14	1	9	147
	6	9	3	8	15	2	10	18	19	7	3	5	99
	7	5	2	6	13	2	8	16	8	7	1	4	72
	8	0	0	2	7	3	0	11	9	1	0	0	33
	9	2	3	0	5	4	4	9	2	2	0	0	31
	10	0	0	0	4	0	0	1	1	0	0	0	6
	11	0	1	1	4	0	1	3	1	0	0	1	12
	12	0	0	0	1	0	0	1	0	1	0	0	3
	13	0	0	0	1	0	0	0	0	0	0	0	1
	14	0	0	0	1	0	0	0	0	0	0	0	1
Second.	0	83	13	59	68	22	36	55	29	56	123	76	620
	1	26	6	17	18	0	11	13	30	18	38	33	210
	2	24	4	3	12	3	3	11	17	18	35	20	150
	3	4	2	1	7	4	1	8	10	12	16	7	72
	4	2	1	1	4	1	4	1	13	9	9	9	54
	5	1	0	0	2	1	0	4	9	4	1	1	23
	6	1	0	0	0	0	1	2	5	0	1	0	10
	7	1	1	0	0	1	1	1	0	0	1	1	7
	9	0	0	0	0	1	0	0	1	0	0	1	3
Higher	0	33	4	24	14	6	27	19	15	81	130	68	421

	1	17	4	9	7	1	12	5	12	18	29	24	138
	2	4	2	6	2	0	11	3	7	25	32	22	114
	3	5	1	1	0	2	1	3	11	7	14	7	52
	4	1	0	0	0	0	3	0	3	0	5	3	15
	5	0	0	0	0	0	0	0	6	0	1	0	7
	6	0	0	0	0	0	1	0	0	0	0	0	1
	7	1	0	0	0	0	0	0	1	0	0	0	2
	8	0	0	0	0	0	0	0	1	0	0	0	1
Total	0	229	141	305	344	210	224	318	194	287	444	343	3039
	1	118	88	120	113	38	94	82	113	108	143	129	1146
	2	88	95	114	115	50	81	96	81	107	117	103	1047
	3	70	86	86	70	58	69	103	77	73	67	55	814
	4	63	60	96	88	47	72	83	69	57	25	46	706
	5	58	46	79	74	50	53	85	65	43	10	44	607
	6	42	38	53	73	57	52	81	59	31	8	23	517
	7	36	39	44	66	48	41	58	29	24	2	20	407
	8	14	21	25	36	21	31	49	23	14	2	16	252
	9	8	11	13	30	31	13	30	8	7	0	14	165
	10	4	7	5	19	13	8	13	1	9	0	9	88
	11	1	6	3	15	11	5	5	3	0	0	7	56
	12	1	1	5	6	2	4	3	1	2	0	1	26
	13	0	2	0	1	2	0	0	0	0	0	1	6
	14	1	0	0	2	1	0	2	0	0	0	1	7
	15	0	0	0	0	1	0	0	0	1	0	0	2
	Total	733	641	948	1052	640	747	1008	723	763	818	812	8885
	Households	171	155	176	167	181	153	166	157	178	181	192	1877

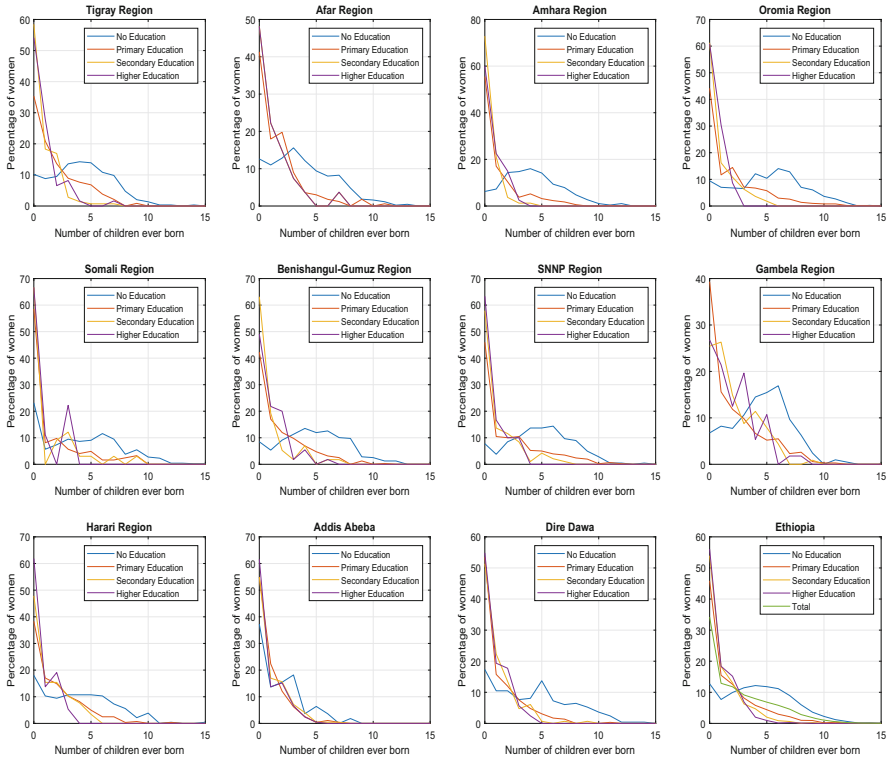


Fig. 1 Children ever born by mother's education and across regions: Ethiopia Mini-DHS 2019

levels. Figures 1 and 2 clearly show that women with no education are over-represented among those with a higher number of children, while women with higher education (above secondary) are over-represented among those with few children (including childless). We also note that the educational differentials in the number of children are not uniform across the regions.

3.2 Sequential Probit Model for Propensities of Parity Progression

Following Ghilagaber and Peristera (2014), let a_i, $i = 0, \ldots, J$ denote the possible number of children, and let, as before, the binary indicator of decision/progress be denoted by W_i. If $w_1 = 0$, outcome a_0 is observed. Otherwise depending on the value of w_2, there are two different outcomes: $\alpha_1 = \{w_1 = 1, w_2 = 0\}$ and $\alpha_2 = \{w_1 = 1, w_2 = 1\}$. Subsequent outcomes can be obtained in a similar way until the outcome α_J. Therefore, the process of parity progression can be viewed

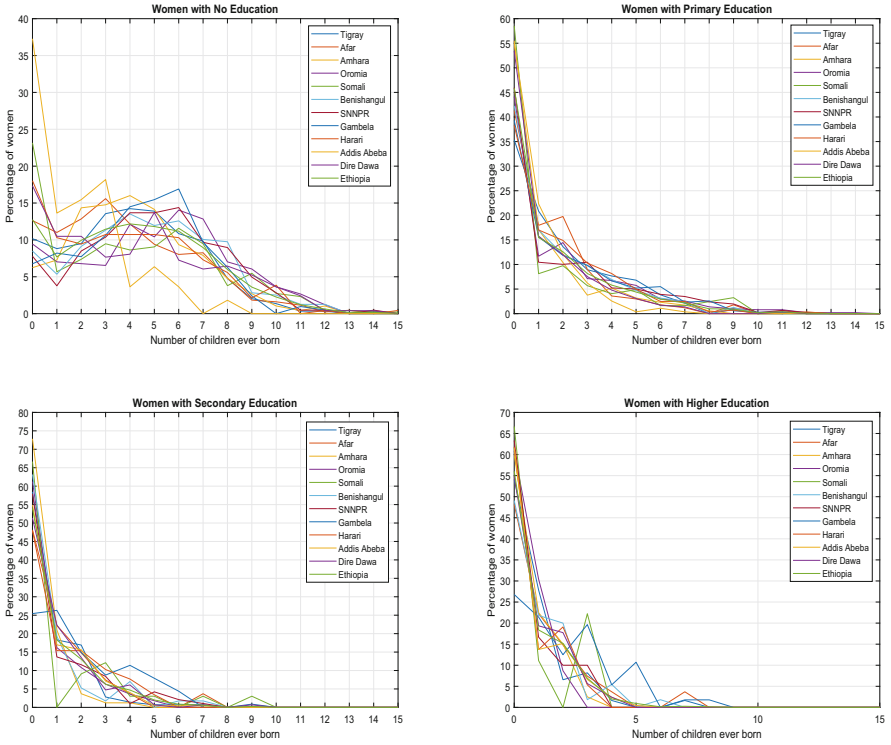


Fig. 2 Children ever born by region and across mother's education: Ethiopia Mini-DHS 2019

as a series of binary choices. See Ghilagaber and Peristera (2014) for a graphical description of the process.

Parity progression process requires successful completion of the prior parity for progress to the next higher parity. Therefore, as argued by Brien and Lillard (1994) and Upchurch et al. (2002), a sequential probit model that assumes parity progression occurs for the J options in a sequential manner accurately reflects the real progression process.

The proposed model specifies an index I_s for the probability of progressing to successively higher parity s, conditional on having completed the previous lower parity. Thus, there are up to J sequential choices of whether to continue to the next parity $(s = 1, \ldots, J)$, each conditional on having completed the previous lower parity. Consequently, if we denote total sample size by N, then $N = \sum n_i,\ i = 1, \ldots, J$ where n_i are the sub-samples of women available at decision level i.

Woman j progresses from parity s to parity $s + 1$ if her propensity to continue is positive, $I_s > 0$. The probability of progression is determined by the probit index function

$$I_s = \alpha_{0s} + \acute{\alpha}_{1s} X_s + \varepsilon^s + t_s \ for\ s = 1, 2, \ldots, J, \tag{7}$$

where X_s is a vector of exogenous covariates affecting parity progression decisions, α_{0s} and α_{1s} are decision-specific intercepts and coefficients, respectively, ε^s is an individual-specific residual term (heterogeneity) affecting all levels of decision, and t_s is the decision-specific stochastic element (normalized to $\sigma_{ts} = 1$, for all s). ε^s and t_s are assumed to be normally distributed:

$$\varepsilon^s \sim N(0, \sigma^2_{\varepsilon s}) \text{ and } t_s \sim iidN(0, 1). \tag{8}$$

The residual terms are assumed independent of each other and all exogenous covariates X_s. The model allows parameters to vary across decisions (hence the subscript s on the parameter vector α). In other words, the parameters α_{1s} can be estimated by dividing the entire sample into smaller sub-samples. The model also allows for correlation between the individual component ε^s and any endogenous explanatory variables.

The probability of any given level of completed parity, s, conditional on the sequence of covariates X is given by

$$P\left[s \mid \Xi(s)\right] = \begin{cases} \int_{\varepsilon^s} f_n\left(\varepsilon^s \mid \sigma^2_{\varepsilon^s}\right) \Phi\left[-\underset{-}{\alpha_s'} X_s + \varepsilon^s\right] \prod_{l=1}^{s-1} \left(\underset{\equiv}{\alpha_l'} X_l + \varepsilon^s\right) d\varepsilon^s, \\ \qquad s = 0, 1, \ldots, J-1 \\ \int_{\varepsilon^s} f_n\left(\varepsilon^s \mid \sigma^2_{\varepsilon^s}\right) \prod_{l=1}^{s-1} \left(\underset{\equiv}{\alpha_l'} X_l + \varepsilon^s\right) d\varepsilon^s, \; s = J \end{cases}, \tag{9}$$

where J is the highest number of children, $\Xi(s)$ denotes the full set of covariates at each of the decision points, $\underset{-}{X_l}$ is the stacked vector of all covariates at each decision l, and $f_n(.)$ is the normal density function.

3.3 Results

Relative propensities of parity progression among all women (for the entire country) are shown in Table 2. The left panel presents results from two probit models where the covariates education and region were entered separately, while the right panel presents results from one multivariate model where both covariates were entered into the model.

The first columns in each panel show results from a model where the effects of the covariates are assumed to be constant across the various parities, while the results in the next five columns of each panel are from models where we let the effects of the covariates to vary across the parities.

The upper part in both panels shows higher propensity of parity progression among women with no education or elementary level education relative to women

Table 2 Relative propensities of progression into various parities by education and region: Ethiopia Mini-DHS 2019[a]

	Results from two univariate models						Results from one multivariate model					
	Progression from and to parity						Progression from and to parity					
Covariate	All parities	0 to 1+	1 to 2+	2 to 3+	3 to 4+	4 to 5+	All parities	0 to 1+	1 to 2+	2 to 3+	3 to 4+	4 to 5+
No Educ	**3.68**	**3.62**	**3.30**	**4.22**	**4.33**	**3.06**	**4.31**	**4.12**	**3.45**	**4.80**	**4.96**	**3.41**
Primary	**1.29**	**1.22**	**1.41**	**1.99**	**2.58**	**2.09**	**1.22**	**1.22**	**1.33**	**1.90**	**2.41**	**1.92**
Secondary	1.05	1.03	1.04	**1.37**	**1.91**	1.14	1.02	1.00	0.98	**1.29**	**1.77**	1.02
Higher	1	1	1	1	1	1	1	1	1	1	1	1
Tigray	1	1	1	1	1	1	1	1	1	1	1	1
Afar	**1.33**	**1.83**	**1.73**	**1.41**	1.28	**1.52**	0.92	1.03	1.08	.89	.87	1.11
Amhara	0.97	0.85	1.04	0.92	0.99	0.87	0.81	0.84	1.01	.84	.97	.89
Oromia	0.96	**0.79**	1.12	0.99	1.27	1.14	0.92	1.00	**1.41**	1.27	**1.65**	**1.47**
Somali	0.96	1.12	**2.34**	**1.83**	**1.53**	**1.75**	**0.55**	**0.52**	1.27	1.06	0.99	1.24
Benishangul	1.04	**1.40**	**1.66**	**1.60**	**1.58**	**1.48**	0.97	0.78	0.95	0.91	0.91	0.86
SNNPR	0.99	**0.80**	**1.34**	1.12	1.00	1.17	0.9	**0.65**	1.14	0.93	0.85	0.98
Gambela	1.14	1.18	1.12	1.18	1.08	1.10	1.27	1.00	1.00	1.11	1.01	1.00
Harari	**0.84**	0.82	1.01	0.80	0.86	0.90	0.88	0.80	1.03	0.85	0.87	0.90
Addis Ababa	**0.55**	**0.56**	**0.64**	**0.43**	**0.35**	**0.45**	**0.66**	**0.58**	**0.72**	**0.55**	**0.41**	**0.53**
Dire Dawa	**0.75**	**0.77**	0.89	0.79	1.01	1.05	0.79	**0.73**	0.88	0.79	0.98	1.02

[a] Estimates that are significant at 5% level (with p-value < 0.05) are shown in **bold** font

with higher education (which is the baseline level). Women with secondary-level education do not differ significantly from those with higher education in most of the columns, but we note that they have significantly higher propensities of progression from parities 2 to 3 and from 3 to 4 in both panels.

In the model with only education and where the effects of education are assumed to be constant over the parities, we found a significant household random effect (p-value = 0.000) that is not reported in the table. However, this effect disappeared once region was included in the model. This should not be surprising because the households belong to the regions, and hence, any household random-effect term is captured by the covariate region. It is also in accordance with Ghilagaber et al. (2022) where household random effects were insignificant in most models for transition to parenthood among Ethiopian women based on data from the 2016 Ethiopian Demographic and Health Survey.

The lower panel of Table 2 shows relative propensities of parity progression by region. The first region (Tigray) was used as a baseline, and we see from the first column that women from the Afar region have significantly higher propensity of parity progression than those in the Tigray region, while women from the Harari region and the two city administrations Addis Ababa and Dire Dawa have significantly lower propensity. Women in the other six regions do not differ significantly from those in Tigray. The corresponding column in the right panel (where education and region are entered in the model) shows that only the Somali region and the capital, Addis Ababa, differ significantly from the Tigray region and that women from these two areas have lower propensity than women from Tigray.

Thus, we already see that we reach at different conclusions with regard to the effect of region on parity progression depending on whether it is included in the model alone or together with education. The same is true when we allow the effects of covariates to vary over the parities. For instance, women from the Harari region have a significant 16% lower propensity in the model where effects are assumed to be constant, while there is no any significant difference in the five models where we allow the effects to vary over parities. The only region that consistently shows significant lower propensities of parity progression in all models is the capital city, Addis Ababa.

These results have, therefore, prompted us to fit a sequential probit model separately for women in each of the 11 regions with only education as a covariate and allowing its effects to vary over the first three progressions. Results from 11 such models are shown in Table 3.

Table 3 shows that women with no education have significantly higher propensities of parity progression compared to those with higher education. For progression from parity 0 to parity 1, this significant difference is true for all regions except the Afar region. Educational differences in the Afar region are insignificant except for progression from parity 1 to parity 2 where women with no education have higher propensities than those with higher education.

In some regions, even women with primary education have significantly higher propensities than the baseline group of women. But, such significant differences vary across the regions and the parities. In some regions (Tigray, Gambela, and

Table 3 Regional relative propensities of progression into parities 1, 2, and 3 by education: Ethiopia Mini-DHS 2019[a]

Parity	Covariate	Tigray	Afar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis	Dire
1st	No Educ.	**3.95**	2.67	**7.18**	**7.19**	**3.49**	**4.25**	**7.41**	**2.67**	**3.93**	**1.95**	**3.18**
	Primary	**1.62**	0.95	1.11	**1.94**	1.24	1.15	**1.79**	**0.68**	**1.92**	1.14	0.98
	Secondary	0.9	0.79	0.65	1.16	1.01	0.66	1.28	1.07	1.31	1.23	1.05
	Higher	1	1	1	1	1	1	1	1	1	1	1
2nd	No Educ.	**4.78**	**4.75**	**6.31**	**16.43**	4.1	**4.8**	**6.45**	**2.45**	**2.67**	**1.64**	**3.29**
	Primary	**2.08**	2.24	1.63	**7.89**	1.92	1.49	**2.36**	1.07	1.33	**0.66**	1.3
	Secondary	1.52	1.37	0.45	**3.83**	–	0.72	1.49	0.87	1.04	0.94	0.91
	Higher	1	1	1	1	1	1	1	1	1	1	1
3rd	No Educ.	**2.32**	5.39	**11.1**	–	–	**7.72**	**3.95**	**2.18**	**10.47**	**2.38**	**7.09**
	Primary	1.16	1.78	**4.47**	–	–	**3.27**	1.94	0.9	**4.34**	1.22	**2.48**
	Secondary	**0.39**	1.4	1.88	–	–	2.75	1.09	0.9	**2.69**	1.12	1.68
	Higher	1	1	1	1	1	1	1	1	1	1	1

[a] Estimates that are significant at 5% level (with p-value < 0.05) are shown in **bold** font

Addis Ababa), we see results that seem contrary to expectations. In Tigray, we see that women with secondary-level education have significantly lower propensity of progression from party 2 to parity 3 than women with higher education. In Gambela, women with primary-level education have significantly lower propensity of progression to parenthood (from parity 0 to parity 1) than those with higher education. In the capital, Addis Ababa, women with primary-level education have significantly lower propensity of progression from parity 1 to parity 2 than those with higher education.

4 Summary and Concluding Remarks

This chapter presented a sequential probit model for parity progression ratios and applied it to data on Ethiopian women surveyed in the 2019 Mini Demographic and Health Survey in the country 2019.

The proposed sequential probit model specifies the propensity of progressing to a successively higher number of children conditional on having attained the current number of children. The model is based on the tacit assumption that a mother's decision at given birth consists of some sequential and independent choices. We argue that the proposed approach captures the decision process on family size more accurately since it conditions the propensity to progress to a given parity on successful completion of the previous lower parity and allows covariate effects to vary across parities.

Our contribution was mainly methodological—to suggest a more appropriate method to analyze the data at hand. However, we also addressed a substantive research question—the effect of woman's educational level on parity progression and how this effect varies across progressions and between the 11 regions in the country.

Our empirical results reveal differentials in parity progression by educational levels and across the regions. But, the strength (significance) and, in some cases, even the direction of educational gradients on parity progression were not uniform across parities and between the regions.

Overall, mothers with higher education have lower propensity of parity progression (specially at higher birth orders). Women from the capital city, Addis Ababa, have much lower propensities to get more children, while women from the Afar and Benishangul–Gumuz regions have much higher propensities of parity progression.

The sequential model proposed in this chapter provides more insight in fertility decision process than conventional probit or logistic regression, and we recommend its use in situations where the response variable is sequentially ordered.

The 8885 women analyzed in this chapter have contributed a total of 23007 children by the survey date (March–June 2019)—an average of 2.59 children per woman. Among these, 22446 (97.6%) were single births, 277 (1.2%) were 1st of multiple births, 277 (1.2%) were 2nd of multiple births, and only 7 (0.03%) were 3rd of multiple births. Since the total of 561 multiple births was only 2.44% of

the total of 23007 births, we have not accounted for twins or multiple births in our analyses. Instead, each birth was treated as unique.

The model we have used in this chapter assumes the error terms in the model are a random sample from a standard normal distribution. It is not guaranteed that this assumption is valid for our data, but diagnostics on this assumption is beyond the scope of this chapter as the aim was to describe and illustrate a modeling approach that we argue is appropriate to the type of data we analyze.

Thus, addressing multiple births (especially when they are an appreciable portion of the total births) and sensitivity of estimates to the choice of distribution of the error terms can be possible topics for future research in the area.

Acknowledgments The data analyzed in this chapter was generously provided by the DHS program (https://www.dhsprogram.com). The views expressed in the chapter are solely of the authors and do not express the views or opinions of the data source or its employees.

Appendix: aML Code Used for Computing the Results in the Upper-Left Panel of Table 2 in This Chapter (Except Those in the First Column)

```
*****************************************************************
 Reading the source file in txt format and obtaining its dat
 version used as input by aML.
-----------------------------------------------------------

ascii data file = Eth2019.txt;
output data file = Eth2019.dat (replace = yes);
level 1 var = Region;
level 2 var = Resid Cohort Educ Children;

*****************************************************************

*****************************************************************
## Fitting a sequential probit model for parity progression with
only Education as a covariate and a household random-effect term.
Note that the effects of education are allowed to vary across five
parity progressions: progression from 0 child to to 1 child (where
all women are included), progression from 1 child to 2 children
(where women with no child are excluded), progression from 2
children to 3 children (where women with no child and those with
only one child are excluded), etc.

aML enables to estimate these set parameters in one run (from the
same model) without having to fit separate models for each
progression.

More details on the program and examples can be obtained in its
website:http://applied-ml.com/
```

```
-----------------------------------------------------------

dsn = Eth2019.dat;

option maximum scratch data space = 9000;
option maximum number of residual draws = 900;

define regset BetaX1;
var = 1 (Educ==0) (Educ==1) (Educ==2);

define regset BetaX2;
var = 1 (Educ==0) (Educ==1) (Educ==2);

define regset BetaX3;
var = 1 (Educ==0) (Educ==1) (Educ==2);

define regset BetaX4;
var = 1 (Educ==0) (Educ==1) (Educ==2);

define regset BetaX5;
var = 1 (Educ==0) (Educ==1) (Educ==2);

define normal distribution; dim=1;
number of integration points=6; name=eta;

/* Model for 0->1*/
probit model; keep if Children>=0; outcome = (Children>=1);
model = regset BetaX1 + intres(draw=_id, ref = eta);

/* Model for 1->2*/
probit model; keep if Children>=1; outcome = (Children>=2);
model = regset BetaX2 + intres(draw=_id, ref = eta);

/* Model for 2->3*/
probit model; keep if Children>=2; outcome = (Children>=3);
model = regset BetaX3 + intres(draw=_id, ref = eta);

/* Model for 3->4*/
probit model; keep if Children>=3; outcome = (Children>=4);
model = regset BetaX4 + intres(draw=_id, ref = eta);

/* Model for 4->5 or higher*/
probit model; keep if Children>=4; outcome = (Children>=5);
model = regset BetaX5 + intres(draw=_id, ref = eta);

starting values;
Cons1       T   -.97076734849
NoEduc1     T    1.2866873162
Prim1       T    .19509099954
Secon1      T    0.0291782222
Cons2       T   -0.6871056844
NoEduc2     T    1.1932315868
Prim2       T    0.3400778214
Secon2      T    .04238392715
```

```
Cons3       T   -1.1742768039
NoEduc3     T    1.4402496565
Prim3       T    .68904298508
Secon3      T    .31155824768
Cons4       T   -1.4092236094
NoEduc4     T    1.4662051764
Prim4       T    .94711322171
Secon4      T    .64751990757
Cons5       T   -1.2394722963
NoEduc5     T    1.1154687783
Prim5       T    .73525501381
Secon5      T      0.12890475
Sdeta       T       0.7
;
```

References

Albert, J. (2009). *Bayesian computation with R*. Springer.

Albert, J., & Chib, S. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics, 57*(3), 829–836.

Alpu, O., & Fidan, H. (2004). Sequential probit model for infant mortality modelling in Turkey. *JApSc, 4*(4), 590–595.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica: Journal of the Econometric Society, 46*(5), 1193–1205.

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press.

Arroyo, J., Payne, K., Brown, S., & Manning, W. (2017). Crossover in median age at first marriage and first birth: Thirty-five years of change. *Family Profile* (22) (accessed online on 20 June 2021).

Brien, M., & Lillard, L. (1994), Education, marriage, and first conception in Malaysia. *Journal of Human Resources, 29*(4), 1167–1204.

Davidson, R., & MacKinnon, J. (2004). *Econometric theory and methods*, Vol. 5. New York: Oxford University Press.

Eickmeyer, K., Payne, K., Brown, S., & Manning, W. (2017). Crossover in the median age at first marriage and first birth: Thirty-five years of change. *Family Profiles*.

Ethiopian-Public-Health-Institute and ICF. (2019). *Ethiopia mini demographic and health survey 2019: Key indicators.* Rockville, Maryland, USA: Ethiopian Public Health Institute and ICF.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models, 2nd edition*. Springer Science and Business Media.

Ghilagaber, G., & Peristera, P. (2014). Sequential probit modelling of family and community effects on educational progress among children to Polish and Turkish immigrants in Sweden. *Quality and Quantity, 48*(6), 3243–3252.

Ghilagaber, G., Gyimah, S., Lofstedt, P., & Johansson, A. (2005). The use of flexible parametric duration functions in modelling the tempo of fertility: applications to the analysis of birth intervals in rural China. *Population Review, 44*(1), 11–35.

Ghilagaber, G., Akinyi Lagehäll, A., & Yemane, E. (2022). Stratified multilevel modelling of survival data: Application to modelling regional differences in transition to parenthood in Ethiopia. In *Modern biostatistical methods for evidence-based global health research*. Springer Series on Emerging Topics in Statistics and Biostatistics.

Gurmu, E., & Etana, D. (2014). Age at first marriage and first birth interval in Ethiopia: analysis of the roles of social and demographic factors. *African Population Studies, 28*(3), 1332–1344.

Liang, Y., & Ghilagaber, G. (2022). Bayesian survival analysis with the extended generalized gamma model: application to demographic and health survey data. In *Modern biostatistical*

methods for evidence-based global health research. Springer Series on Emerging Topics in Statistics and Biostatistics.

Liao, T. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*, vol. 101. Sage.

Lillard, L. A., & Panis, C. W. A. (2003). *aML Multilevel multiprocess statistical software, version 2.0.*. Los Angeles, CA: EconWare.

Maddala, G. (1986). *Limited-dependent and qualitative variables in econometrics*, vol. 3. Cambridge University Press.

Mare, R. (1980). Social background and school continuation decisions. *Journal of the American Statistical Association, 75*(370), 295–305.

Marini, M. (1981). Effects of the timing of marriage and first birth on fertility. *Journal of Marriage and the Family, 43*(1), 27–46.

Munezero, P., & Ghilagaber, G. (2022). Dynamic Bayesian modelling of educational and residential differences in family initiation among Eritrean men and women. In *Modern biostatistical methods for evidence-based global health research*. Springer Series on Emerging Topics in Statistics and Biostatistics.

Munkin, M. (2011). The endogenous sequential probit model: An application to the demand for hospital utilization. *Economics Letters, 112*(2), 182–185.

Nagakura, D. (2004). A note on the relationship of the ordered and sequential probit models to the multinomial probit model. *Economics Bulletin, 3*(40), 1–7.

Steele, F., & Durrant, G. (2011). Alternative approaches to multilevel modelling of survey noncontact and refusal. *International Statistical Review, 79*(1), 70–91.

Upchurch, D., Lillard, L. A., & Panis, C. (2002). Nonmarital childbearing: Influences of education, marriage, and fertility. *Demography, 39*(2), 311–329.

Waelbroeck, P. (2005). Computational issues in the sequential probit model: a Monte Carlo study. *Computational Economics, 26*(2), 141–161.

Wood, J., Neels, K., & Kil, T. (2014). The educational gradient of childlessness and cohort parity progression in 14 low fertility countries. *Demographic Research, 31*, 1365–1416.

Propensity Score Approaches for Estimating Causal Effects of Exposures in Observational Studies

Halima S. Twabi, Samuel O. M. Manda, and Dylan S. Small

Abstract As regards study designs, randomised controlled trials are judged as the gold standard for quantitatively evaluating treatment effect sizes with less bias than observational trials. In some cases, the RCTs can be considered unethical, not feasible and impractical to conduct. In such cases, when RCTs are not appropriate for evaluating interventions, observational studies, which generate valuable health data and are readily available, have been used. A major disadvantage of observational studies is that they cannot be used for investigating cause–effect relationships due to confounding factors. Propensity score approaches are one of the strategies that have been developed to control for confounder bias in observational studies and allow for the estimation of causal association. This chapter provides a description and theoretical fundamentals of two propensity-score-based approaches, namely the propensity score matching and propensity score weighting for facilitating the assessment of causal exposure effects using observational data. The two methods are illustrated with an evaluation of the effects of: (a) exclusive breastfeeding or (b) appropriate complementary feeding on nutritional outcomes of infants or children using survey data from Malawi and Zambia.

H. S. Twabi (✉)
Department of Mathematical Sciences, University of Malawi, Zomba, Malawi
e-mail: htwabi@unima.ac.mw

S. O. M. Manda
Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: samuel.manda@up.ac.za

D. S. Small
Department of Statistics, University of Pennsylvania, Philadelphia, PA, USA

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
e-mail: dsmall@wharton.upenn.edu

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_4

Keywords Causal inference · Propensity score · Propensity score matching · Inverse probability weighting · Observational studies · Treated group · Control group · Exclusive breastfeeding · Complementary feeding · Sensitivity analysis · Confounders · Selection bias · Confounder bias · McNemar's test · Mantel–Haenszel test · Conditional logistic regression · Paired data · Child nutritional outcomes · Child malnutrition · Stunting · Wasting · Underweight · Sub-Saharan Africa · Average treatment effect · Confounder balance · Randomised control trials · Absolute standardised differences · Demographic and Health surveys · Child growth · Ignorability assumption · Model misspecification · Nutritional interventions · Anthropometric measures · Maternal HIV · Child growth measurements · Positivity · Consistency · Exchangeability

1 Introduction

An important objective of empirical comparison studies in health research is the estimation of causal effects of a treatment, exposure, or intervention on health outcomes. Ideally, randomised control trials (RCTs) are the gold standard design to assess causal effects (Rubin, 1973). RCTs ensure a random allocation of subjects into treated and control groups. In this case, treated subjects do not differ systematically from control subjects in both measured and unmeasured baseline characteristics. Therefore, this renders the possibility to directly estimate the effect of a treatment by comparing the outcomes between the treated and control groups. However, sometimes RCTs can be considered unethical and impractical to conduct. The availability of observational studies, which are rich with valuable data, has enabled health researchers to estimate the causal effect of an exposure on an outcome. However, observational studies render an assessment of causal association not possible.

Non-randomised studies of the effect of treatment on outcomes can be subject to bias in which treated subjects differ systematically from control subjects (Rosenbaum & Rubin, 1983). For example in health research, a patient may be given a treatment, and a clinical researcher would observe the outcome, but the treatment allocation in this situation is not random. On the other hand, in health survey data, there is a lack of randomisation and treatment assignment, and both the exposure and the outcome are observed. Therefore, the estimation of causal effects in the former scenario may be subjected to selection bias, while for the latter example, the effects may be affected by confounder bias. Estimation of the treatment effect cannot be done by simply comparing outcomes between treatment groups. The causal effect of interest would be misleading due to the presence of confounders. Such confounding can result in biased estimates because confounding essentially means that some causes of the outcome also influence selection for the exposure (Rosenbaum & Rubin, 1984; Abadie et al., 2004; Austin & Mamdani, 2006). Even though treatment assignment is mentioned when discussing causal effects, however,

the assessment of causal association is not limited to a health setting. Different areas that require an assessment of causal effects such as intervention effectiveness or exposure effects can be done. In this chapter, we use the term treatment and exposure interchangeably.

Various strategies have been developed and used to address confounder bias in non-randomised treatment comparison studies such as control for covariates in the analysis through multivariate regression (Kurth et al., 2006). Ideally, the aim is to identify all confounders that influence the exposure and outcome, and then the differences found between the treatment and control groups after correctly adjusting for the identified covariates will represent the causal effects. Other strategies involve matching the treated and control subjects based on similar observed covariates (Rubin, 1973; Abadie et al., 2004). For example, if x_i denoted the set of observed covariates age, gender, and eating habits (healthy or unhealthy), each treated subject is paired with a control subject, having the same gender, same (or similar) age, and eating habits. However, as the dimensionality of the covariates increases, matching subjects with respect to a large number of covariates tends to be difficult. The introduction of propensity score (PS) methods by Rosenbaum and Rubin (1983) has enabled an easy direction in the control for confounding effects in observational studies when comparing treatment effects.

The propensity score (PS) is the conditional probability of being treated given observed covariates (Rosenbaum & Rubin, 1983). This implies that conditional on the measured baseline covariates, allocation of subjects to treatment groups is considered to be a random process that mimics RCTs (Austin, 2011a). This is possible because using the PS, observations in the treated and control groups with similar PS have nearly similar observed distributions of covariates (d'Agostino, 1998) and are comparable. A comparison of the outcomes between the two groups would represent the causal effect. The most widely used PS methods are the propensity score matching (PSM) (Rosenbaum & Rubin, 1983, 1985) and inverse probability weighting (IPW) (Robins et al., 1994) on the PS. Both the PSM and IPW use the propensity scores to create a sample of control and treated subjects that have similar characteristics. The propensity score matching (PSM) involves matching the subjects based on the (estimated) propensity scores (Rosenbaum & Rubin, 1983). Matching on the PS results in the analysis based upon only those subjects who are successfully matched. The inverse probability of treatment involves assigning a weight to a subject based on the propensity score. The re-weighted subjects in the treated and control groups create a pseudo-population in which there is no association between confounders and treatment. The advantage of the IPW is that all observations are included in the weighted population unlike for the PS matching where the treatment effect is estimated on the matched sample (Lunceford & Davidian, 2004; Cole & Hernán, 2008; Austin, 2011a). One of the challenges on weighting on the PS is that exposed subjects with a very low propensity score can result in a very large weight. Similarly, a control subject with a propensity score close to one can result in a very small weight.

The purpose of this chapter is to provide statistical descriptions and a theoretical background of the propensity score matching and inverse probability weighting

methods that are used to correct for confounder biases in observational studies, allowing for causal-effect treatment comparisons. Their usage is demonstrated in the evaluation of the effect of: (a) exclusive breastfeeding or (b) appropriate complementary feeding on improving child growth. Child growth indicators are taken to be binary as well as continuous.

1.1 Infant and Young Child Feeding Interventions and Child Nutritional Outcomes

There is growing evidence of the critical consequences of exclusive breastfeeding (EBF) in the first six months of life on child health and nutritional outcomes (WHO et al., 2009a; Kuchenbecker et al., 2015; Kumar & Singh, 2015; Ayisi & Wakoli, 2014; Kamenju et al., 2017; Perkins et al., 2018). Exclusive breastfeeding entails feeding an infant only breast milk with an exception of drops or syrups consisting of vitamins, mineral supplements, or medicine. Breast milk contains essential nutrients for a child, such as vitamins and minerals, which have been found to be protective against common childhood illnesses such as gastrointestinal infections and pneumonia (WHO, 2003; WHO et al., 2009b). However, adequate nutrition can be provided only to a certain age at which point a child needs additional solid foods to supplement for the deficiency. The World Health Organisation recommends introducing complementary semi-solid, soft foods and solid foods, liquids, water along with breast milk to children at the age of 6 months (WHO et al., 2009b). Appropriate complementary feeding, which involves initiating complementary feeding (nutrient-rich foods) at the appropriate time and feeding a child sufficiently, has been shown to have an impact on child nutritional outcomes among children aged between 6 and 23 months (WHO et al., 2009b; Kassa et al., 2016). Poor feeding habits among infants and young children are associated with increased risk of illnesses, frequency of infections, and reduced nutrition absorption that result in poor growth (WHO, 2003; WHO et al., 2009b). Optimal infant and young child feeding is known to be essential in ensuring good child growth and health (WHO et al., 2009b). Studies that have provided evidence of the beneficial effect of exclusive breastfeeding and appropriate complementary feeding on child nutritional outcomes have based their conclusions from observational studies that are prone to confounder bias. Malawi and Zambia are among the countries in SSA that have been adversely affected by poor child growth, particularly stunting among under-five-year-old children (WHO et al., 2017). There is a need to affirm the evidence of the beneficial effect of infant and young child feeding interventions in improving child growth using advanced statistical methods. In this chapter, we assessed three nutritional outcomes, namely: stunting (height-for-age), wasting (weight-for-height), and underweight (weight-for-age). The three nutritional outcomes are standardised to *z-scores* such that values below -2 indicate adverse outcomes (WHO et al., 2006). In the case of height-for-age *z-score*, a child with a score below -2 is classified as stunted, for weight-for-age

z-score, a child with a score below −2 is classified as underweight, and for weight-for-height *z-score*, a child with a score below −2 is classified as wasted.

2 PS Methods to Minimise Confounder Bias in Estimating Exposure Effects

Let T be a binary treatment indicator denoted as $T = \{0, 1\}$, where $T = 1$ if treated (e.g., exclusively breastfed), $T = 0$ if control (not exclusively breastfed), and $\mathbf{X}$ be a vector of measured covariates that are thought to be associated with the treatment (e.g., exclusive breastfeeding) and the outcome (e.g., child height-for-age z-score). Each subject (i.e., a child) is assumed to have a vector of hypothetical outcomes $Y_T = (Y_0, Y_1)$. The Y_0 and Y_1 represent values of the height-for-age z-score that would be observed to have a child been treated or have the child received a control. These hypothetical outcomes are known as potential outcomes (or counterfactuals). The counterfactual outcomes Y_0, Y_1 represent a hypothetical situation where at a population level, all children are not exclusively breastfed and when all children are exclusively breastfed. The actual observed outcome of a child when exposed is denoted as

$$Y = Y_1 T + (1 - T) Y_0. \tag{1}$$

Equation 1 is also referred to as the consistency assumption, where $Y = Y_T$ (Rosenbaum & Rubin, 1983; Austin & Mamdani, 2006). This assumption states that the observed outcome is equal to the counterfactual under the actual treatment a subject received. Due to the *fundamental missing data problem* in causal inference (Holland, 1986), an average causal effect is estimated from the population rather than individual causal effects. The difference in the mean potential outcomes is denoted as

$$\begin{aligned} &E[Y_1 - Y_0] \\ &E[Y_1] - E[Y_0]. \end{aligned} \tag{2}$$

The estimation of Eq. 2 cannot be realised from observational studies as the counterfactuals are both not observed from an individual. However, it is possible to identify the causal association by estimating the mean counterfactual outcomes from observed data $O = (Y, T, X)$ under several assumptions known as the *strongly ignorability* assumption.

Under randomised assignment of subjects into treatment groups, the counterfactual outcomes and treatment assignment are independent $Y_1, Y_0 \amalg T$; hence, using observed data, we can show that the expectation of the observed outcome given a

particular treatment is equal to the expected counterfactual outcome at that treatment level $E[Y|T = 1] = E[Y_1|T = 1] = E[Y_1]$. However, for observational data, the presence of characteristics X that influence the causal relationship results in

$$E[Y|T = 1] = E[Y_1|T = 1] \neq E[Y_1].$$

Identifying these covariates and controlling for them may ensure an identification of the average causal effect from the observed data. Therefore, Rosenbaum and Rubin (1983) improved the independence assumption to condition on observed covariates written as

$$(Y_1, Y_0) \amalg T|\mathbf{X}, \tag{3}$$

where T is the observed exposure and $\mathbf{X}$ is a vector of observed covariates. Assumption 3 together with the positivity assumption $0 < p(T = 1|\mathbf{X}) < \mathbf{1}$, which states that each subject within a population is equally likely to be treated, is known as the *strongly ignorability* assumption (Rosenbaum & Rubin, 1983). The assumption states that conditional on observed covariates, exposure is independent of the potential outcomes and, hence, necessitates exchangeability between exposure groups. Using this assumption, we can prove that the mean of the potential outcomes can be estimated from observational data as follows:

$$\begin{aligned} E\{E(Y|T = 1, \mathbf{X})\} &= E\{E(Y_1|T = 1, \mathbf{X})\} \\ &= E\{E(Y_1|\mathbf{X})\} \\ &= E(Y_1). \end{aligned} \tag{4}$$

The first expression of the equation is due to iteration of expectation and consistency. The second expression is due to the strongly ignorable assumption. The same can be shown for $E\{E(Y|T = 0, \mathbf{X})\}$.

2.1 Propensity Score Definition

Rosenbaum and Rubin (1983) introduced an alternative to the traditional balancing method between exposed and unexposed subjects on covariates called the propensity score. The propensity score is defined as the probability that a subject is assigned to a treatment group given observed covariates $\mathbf{X}$.

$$\pi(x) = Pr(T = 1|\mathbf{X}) \tag{5}$$

$0 < \pi(x) < 1$.

The propensity score provides a single measure of influence of confounders on exposure assignment. Equation 5 states that T and X are independent conditional on the propensity score $X \perp\!\!\!\perp T|\pi(x)$. This allows those subjects from the treated and control groups with the same propensity score to be balanced with respect to the distribution of $\mathbf{X}$. Rosenbaum and Rubin (1983) proved that the propensity score is a balancing score, and conditioning on the PS, the potential outcomes are independent of exposure for $\pi\varepsilon(0, 1)$. To prove the balancing property of the propensity score $P(T = 1|\pi(x), x) = \pi(x)$,

$$\begin{aligned} P(T = 1|\pi(x)) &= E(T|\pi(x)) \\ &= E[E(T|\pi(x), x)|\pi(x)] \\ &= E[E(T|x)|\pi(x)] \\ &= E[Pr(T = 1|x)|\pi(x)] \\ &= E[\pi(x)|\pi(x)] \\ &= \pi(x). \end{aligned} \tag{6}$$

Hence, we can further define assumption 3 as $(Y_1, Y_0) \perp\!\!\!\perp T|\pi(x)$ and show that conditional on the propensity score, the mean potential outcomes may be identified from the observed data.

$$E\{E[Y|T = 1, \pi(x)]\} = E\{E[Y_1|T = 1, \pi(x)]\} = E\{E(Y_1|\pi(x))\} = E(Y_1).$$

The same applies for $T = 0$, resulting in $E\{E[Y|T = 1, \pi(x)]\} - E\{E[Y|T = 0, \pi(x)]\} = E[Y_1 - Y_0]$

2.1.1 Propensity Score Estimation

The common approach used to estimate the propensity score is assuming a logistic distribution for the PS and estimating the probability using a logistic regression model. The estimated propensity score is the predicted probability of the exposure from the fitted regression model (Austin, 2011a) written as

$$logit(\pi(x_i)) = log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \log\left(\frac{Pr(T_i = 1|x_i)}{1 - Pr(T_i = 1|x_i)}\right) = \mathbf{X}_i\boldsymbol{\beta}.$$

In multivariate form, this can be written as $\pi(\mathbf{X}, \boldsymbol{\beta}) = \{1 + exp(-\mathbf{X}'\boldsymbol{\beta})\}^{-1}$, and $\boldsymbol{\beta}$ is a $(p \times 1)$ matrix of coefficients. Interactions and higher-order terms can be included

in the model. The parameter $\boldsymbol{\beta}$ is estimated by the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\beta}}$ solving (Lunceford & Davidian, 2004);

$$\sum_{i=1}^{n} \psi_{\beta}(T_i, \mathbf{X}_i, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{T_i - \pi(\mathbf{X}_i, \boldsymbol{\beta})}{\pi(\mathbf{X}_i, \boldsymbol{\beta})\{1 - \pi(\mathbf{X}_i, \boldsymbol{\beta})\}} \frac{\partial}{\partial \beta}\{\pi(\mathbf{X}_i, \boldsymbol{\beta})\} = \mathbf{0}. \tag{7}$$

Although estimation of the propensity scores is common using the logistic regression, studies have estimated the p-scores using probit models, boosted regression methods (McCaffrey et al., 2004), tree-based methods (Lee et al., 2010), and neutral networks (Setoguchi et al., 2008). Variables that are considered as important for matching are included as covariates in the logistic model.

Under the strongly ignorability assumption, using the propensity scores, estimation of an unbiased average treatment effects is possible (Rosenbaum & Rubin, 1983). Methods of propensity score matching, stratification, weighting, and covariate adjustment have been developed to facilitate the causal inference estimation using propensity scores (Austin, 2011a; Cole & Hernán, 2008; d'Agostino, 1998; Rosenbaum & Rubin, 1984; Lunceford & Davidian, 2004).

2.1.2 Propensity Score Model Misspecification

The ideal situation when dealing with propensity scores is to have a known PS. However, in reality, the PS is estimated from the observed data. Some studies have used non-parametric methods to estimate the propensity score (Zhang, 2017). However, the most common method used to estimate the propensity score for a binary exposure is by using a parametric model, mostly, the logistic regression as explained in the previous section. The estimation of the average treatment effect is also mostly done using parametric models. Therefore, the PS model is prone to misspecification, either being under-specified (ignoring interaction or high-order terms) or when a relevant covariate is excluded. Drake (1993) found that misspecification as a result of excluding relevant covariates resulted in an increase in bias but could be resolved by including all necessary covariates. In addition, they observed that misspecifying the PS model resulted in smaller bias as compared to misspecifying the outcome model. Extensions on assessing misspecification of the PS model though limited have been done for complex survey data (Lenis et al., 2018).

2.2 Propensity Score Matching

Traditional matching involves pairing treated and control subjects based on one or several measured covariates. However, matching on the covariates becomes difficult

as the dimensionality of the covariates increases. The introduction of the propensity score by Rosenbaum and Rubin (1983) made it possible to match treated and control subjects based on the estimated propensity score resulting in two groups that are comparable, and hence, the average treatment effect can be estimated from this matched sample (Rosenbaum & Rubin, 1983) using methods for paired tests (paired t-test, McNemar's test, or conditional logistic regression) (Austin, 2009).

Various matching methods have been proposed in the literature to ensure optimal matching, since using the estimated PS alone to match would be limited as the probability of observing two units with exactly the same value of the propensity score is zero (Becker & Ichino, 2002). These methods are the nearest neighbour matching, Calliper matching, Kernel matching, and Mahalanobis metric matching. The nearest neighbour matching is based on the greedy matching algorithm that matched each subject i in the treated group with a subject j in the control group by the smallest absolute distance between their propensity scores:

$$d_i = min_j |\pi(\mathbf{X}_i) - \pi(\mathbf{X}_j)|.$$

Calliper matching pairs each subject i in the treated group with subject j in the control group within a pre-specified calliper region b:

$$d_i = min_j \{|\pi(\mathbf{X}_i) - \pi(\mathbf{X}_j)\}| < b.$$

This helps in reducing the risk of poor matches when the distance of the propensity scores between the matched pairs is great. Mahalanobis metric matching with the propensity score matches each subject i in the treated group with a subject j in the control group according to the closest Mahalanobis distance calculated on the similarities of the variables:

$$d_i = min_j |D_{ij}|,$$

where $D_{ij} = (\mathbf{W}_i - \mathbf{W}_j)\mathbf{S}^{-1}(\mathbf{W}_i - \mathbf{W}_j)^T$, $\mathbf{W}$ is a combined matrix of $\{\mathbf{X}, \pi(\mathbf{X})\}$ and $\mathbf{S}$ is the sample variance–covariance matrix of $\mathbf{X}$ for the control group (Rosenbaum & Rubin, 1985).

Several improvements have been done on the various matching methods. Dehejia and Wahba (2002) introduced the radius matching that is a form of a calliper matching except that the matching is one-to-many with each subject i in the treated group being matched with multiple subjects in the control group within a pre-specified calliper region. Pan and Bai (2015) extended the calliper matching to interval matching that matches subjects based on confidence intervals in propensity scores. Further extensions include the Mahalanobis calliper matching (Guo et al., 2006) and the genetic matching that are forms of the Mahalanobis metric matching that uses callipers and weighted Mahalanobis, respectively.

Several methods have been proposed in the literature to estimate the treatment effects from the matched sample. The simple approach involves comparing the mean of the pair differences written as

$$\widehat{\delta} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{j1} - \bar{Y}_{j0},$$

where j represents the n_1 matched sets and $\bar{Y}_{j0}$ is the average of K control subjects in the matched set j (Austin, 2014). We can define the matching with replacement estimator that attempts to utilise all observations in the data by pairing each subject in the treated or control group with M subjects of the opposite treatment assignment. The matching replacement estimator can be written as

$$\widehat{\delta}_{matchrep} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_{i1} - \widehat{Y}_{i0})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\left(1 + \frac{K_M(i)}{M}\right) Y_i,$$

where

$$\widehat{Y}_{i1} = \begin{cases} \frac{1}{M}\sum_{j\epsilon J_M(i,\phi)} Y_i & \text{if} \quad T_i = 0 \\ Y_i & \text{if} \quad T_i = 1 \end{cases}$$

$$\widehat{Y}_{i0} = \begin{cases} Y_i & \text{if} \quad T_i = 0 \\ \frac{1}{M}\sum_{j\epsilon J_M(i,\phi)} Y_i & \text{if} \quad T_i = 1. \end{cases}$$

$J_M(i,\phi)$ is the set of M observations in the treated group opposite to i with similar propensity scores and $K_M(i)$ is the number of times a subject i is used as a match (Abadie & Imbens, 2006). Abadie and Imbens (2006) proposed an estimator of the variance for the matching with replacement estimator written as

$$\widehat{Var}(\widehat{\delta}_{matchrep}) = \frac{1}{n^2}\sum_{i=1}^{n}(\widehat{Y}_{i1} - \widehat{Y}_{i0} - \widehat{\delta}_{matchrep})^2$$

$$+\frac{1}{n^2}\sum_{i=1}^{n}\left[\left(\frac{K_M(i)}{M}\right) + \left(\frac{2M-1}{M}\right)\left(\frac{K_M(i)}{M}\right)\right]\widehat{\sigma}^2(X_i, T_i), \tag{8}$$

where M is the number of matches per subject, $K_M(i)$ is the number of times subject i is used as a match, and $\widehat{\sigma}^2(X_i, T_i)$ is the estimated conditional variance given by

$$\widehat{\sigma}^2(X_i, T_i) = \frac{J}{J+1}\left(Y_i - \frac{1}{J}\sum_{m=1}^{J} Y_{s_m(i)}\right)^2, \tag{9}$$

where J is the fixed number of similar subjects and $s_m(i)$ is the mth closest subject to subject i among subjects with the same A-value (Abadie & Imbens, 2006).

2.3 *Inverse Probability Weighting (IPW) on the PS*

Another balancing method that reduces confounder bias in observational studies is weighting on the propensity score. The PS weighting method assigns to each subject a weight that equals the inverse of the probability of being treated. Weighting creates a pseudo-population where the distribution characteristics between the treated and control groups are similar (no confounding) (Lunceford & Davidian, 2004). The inverse probability weight is the inverse of the estimated propensity score and is similar to weights used in survey data (Horvitz & Thompson, 1952). A subject in a treated group is assigned a weight of $w_i = \frac{1}{\pi(x_i)}$, and a control subject would be given a weight of $w_i = \frac{1}{1-\pi(x_i)}$. The weighted sample can then be used to estimate the exposure effect on the outcome. The weighted sample represents subjects who have unconfounded estimates. The inverse probability weights can be defined with respect to the exposure as

$$w_i = \frac{T_i}{\pi(x_i)} + \frac{(1-T_i)}{1-\pi(x_i)}, T_i = 0, 1. \tag{10}$$

Therefore, the mean potential outcomes for the treated and control groups can be estimated 1 as $E\{\frac{TY}{\pi(x)}\} = E\{\frac{TY_1}{\pi(x)}\}$. Under the consistency assumption, Y_1 is observed if $T = 1$ and Y_0 is observed if $T = 0$. Therefore, under the strongly ignorability assumption, the potential outcomes can be estimated from the observed data by

$$\begin{aligned} E\left\{\frac{TY}{\pi(x)}\right\} &= E\left\{E\left[\frac{I(T=1)Y_1}{\pi(x)}\middle| Y_1, \mathbf{X}\right]\right\} \\ &= E\left\{\frac{Y_1}{\pi(x)}E[I(T=1)|Y_1, \mathbf{X}]\right\} \\ &= E\left\{\frac{Y_1}{\pi(x)}P[T=1|Y_1, \mathbf{X}]\right\} \end{aligned}$$

$$= E\left\{\frac{Y_1}{\pi(x)} P[T=1|\mathbf{X}]\right\}$$
$$= E\left\{\frac{Y_1}{\pi(x)} * \pi(x)\right\}$$
$$= E[Y_1].$$

Denoting the indicator of the treatment $T = I(T = 1)$. The same applies to $E\{\frac{TY}{\pi(x)}\} = E\{\frac{TY_0}{\pi(x)}\} = E[Y_0]$. Therefore, the average treatment effect can be obtained as follows (Rosenbaum & Rubin, 1985):

$$\widehat{\theta}_{ipw1} = \mu_{1,ipw} - \mu_{0,ipw}$$
$$= E\left[\frac{TY}{\pi(x)} - \frac{(1-T)Y}{\pi(x)}\right]$$
$$= E[\frac{TY}{\pi(x)}] - E[\frac{(1-T)Y}{\pi(x)}]$$
$$= n^{-1}\sum_{i=1}^{n}\frac{T_i Y_i}{\widehat{\pi}} - n^{-1}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-\widehat{\pi}}. \tag{11}$$

The weights in 11 do not add to one; hence, Lunceford and Davidian (2004) proposed a normalised estimator for the average treatment effect where the weights add to one for each group as $E\left[\frac{T}{\pi(\mathbf{X})}\right] = E\left[\frac{1-T}{(1-\pi(\mathbf{X}))}\right] = 1$. Hence, the second version of the IPW estimator is defined as

$$\widehat{\theta}_{ipw2} = \left(\sum_{i=1}^{n}\frac{T_i}{\widehat{\pi}(\mathbf{X})}\right)^{-1}\sum_{i=1}^{n}\frac{T_i Y_i}{\widehat{\pi}(\mathbf{X})} - \left(\sum_{i=1}^{n}\frac{(1-T_i)}{(1-\widehat{\pi}(\mathbf{X}))}\right)^{-1}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-\widehat{\pi}(\mathbf{X})}. \tag{12}$$

$\pi(\mathbf{X})$ is estimated by a logistic regression. The estimators 11 and 12 are solutions to the equations (Lunceford & Davidian, 2004).

$$\sum_{i=1}^{n}\left\{\frac{T_i(Y_i-\mu_1)}{\pi_i}\right\} + \eta_1\left(\frac{T_i-\pi_i}{\pi_i}\right) = 0, \text{ and} \tag{13}$$

$$\sum_{i=1}^{n}\left\{\frac{(1-T_i)(Y_i-\mu_1)}{(1-\pi_i)}\right\} + \eta_0\left(\frac{T_i-\pi_i}{(1-\pi_i)}\right) = 0. \tag{14}$$

With the assumption that the propensity score is known and letting $(\eta_0, \eta_1) = (\mu_0, \mu_1)$ to obtain $\widehat{\theta}_{ipw1}$, while $(\eta_0, \eta_1) = (0, 0)$ to yield $\widehat{\theta}_{ipw2}$.

The variances of the IPW estimators are derived by handling the estimates as solutions to a set of estimating equations as Hernán and Robins (2020) recommended fitting generalised estimating equations using the sandwich variance estimator. Lunceford and Davidian (2004) included the propensity score in the estimating equations, and it is known to be exact. Using the theory of M-estimation that implies that for the IPW estimators $\widehat{\theta}$, $n^{1/2}(\widehat{\theta} - \theta)$ converges in distribution to a $N(0, \Sigma)$ variable. We can show that the large sample variance of $\widehat{\theta}_{ipw1}$ assuming the PS is known is derived as

$$\Sigma^*_{ipw1} = E\left[\frac{(Y^1)^2}{\pi} + \frac{(Y^0)^2}{1-\pi}\right] - \widehat{\theta}_0^2$$

and

$$\Sigma^*_{ipw2} = E\left[\frac{(Y^1 - \mu_1)^2}{\pi} + \frac{(Y^0 - \mu_0)^2}{1-\pi}\right] - \widehat{\theta}_0^2,$$

where $\widehat{\theta}_0$ is the true value, $\mu_1 = E[Y^1]$, $\mu_0 = E[Y^0]$.

Now, assuming that we estimate the propensity scores using the logistic regression model and we estimate $\boldsymbol{\beta}$ by maximum likelihood by solving 7, therefore, using M estimators, the estimator $\widehat{\theta}_{ipw1}$ will be

$$\boldsymbol{\phi}(\mathbf{Y}_i, \widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{(Y^1)^2}{\pi} + \frac{(Y^0)^2}{1-\pi} \\ \frac{T_i - \pi(\mathbf{X}_i, \boldsymbol{\beta})}{\pi(\mathbf{X}_i, \boldsymbol{\beta})\{1-\pi(\mathbf{X}_i, \boldsymbol{\beta})\}} \end{pmatrix}.$$

Therefore, the large sample variances of $\widehat{\theta}_{ipw1}$ and $\widehat{\theta}_{ipw2}$ become

$$\Sigma_{ipw1} = \Sigma^*_{ipw1} - \mathbf{H}^T_{\beta,1}\mathbf{E}^1_{\beta,\beta}\mathbf{H}_{\beta,1},$$

where

$$\mathbf{H}_{\beta,1} = E\left[\left(\frac{Y^1}{\pi} + \frac{Y^0}{1-\pi}\right)\pi_\beta\right]$$

$$\Sigma_{ipw2} = \Sigma^*_{ipw2} - \mathbf{H}^T_{\beta,2}\mathbf{E}^1_{\beta,\beta}\mathbf{H}_{\beta,2},$$

where

$$\mathbf{H}_{\beta,2} = E\left[\left(\frac{Y^1 - \mu_1}{\pi} + \frac{Y^0 - \mu_0}{1-\pi}\right)\pi_\beta\right]$$

and $\pi_\beta = \frac{\partial}{\partial \beta}\pi(\mathbf{X}, \boldsymbol{\beta})$ and $\mathbf{E}_{\beta,\beta} = E\left[\frac{\pi_\beta \pi_\beta^{'}}{\pi(1-\pi)}\right]$. It has been shown that estimating $\boldsymbol{\beta}$ leads to a smaller large sample variance than using the true value of $\boldsymbol{\beta}$ (Lunceford & Davidian, 2004).

2.4 Assessing Confounder Balance

Covariate balance is typically assessed and reported by using a variety of statistical measures such as the standardised mean differences, overlapping coefficient, variance ratios, p-values t-test or Kolmogorov–Smirnov test, and the bias (Austin, 2009). The most common method used to assess covariate balance is the standardised difference. Even though there is no firm agreement on what value of the standardised difference denotes imbalance between treated and control subjects in the matched sample, some researchers proposed that a standardised difference of 0.1 (10%) denotes meaningful balance in the measured covariates (Normand et al., 2001).

To assess the reduction of selection bias associated with a covariate X_p, $p = 1, \ldots, P$ can be calculated by taking the mean difference in the covariates between the treatment groups.

$$B_p = M_1(X_p) - M_0(X_p),$$

where M_1 and M_0 are the means of the covariate for the treatment and control groups.

Alternatively, the standardised bias for a covariate can be used to assess balance. The standardised bias compares the mean and prevalence of the covariate in the treated and control groups written as

$$SB_p = \frac{B_p}{sp_p} = \frac{M_1 - M_0}{sp_p}, \tag{15}$$

where sp_p is the pooled standard deviation of the covariate (Austin, 2009).

For inverse probability weighting, for each covariate (X_p), we calculate a weighted mean or proportion among individuals given an intervention $T_i = 1$ and a weighted mean or proportion among individuals not on an intervention $T_i = 0$ and obtain the difference and divide by the square root of the weighted variance. The weighted mean for individuals on an intervention is calculated as

$$\bar{x}_{wt,T=1} = \frac{\sum I(T_i = 1) w_1 x_i}{\sum I(T_i = 1) w_1},$$

and the weighted sample variance is given as

$$s^2_{wt,T=1} = \frac{\sum w_1}{(\sum w_1)^2 - \sum w_1^2} \sum w_1 (x_i - \bar{x}_{wt})^2$$

and likewise for non-intervention individuals, where w_1 is the weight assigned to the i-th individual on an intervention. A decision criterion using standardised bias varies, but most studies conclude that covariates are balanced if the $|bias| < 5\%$.

2.4.1 Sensitivity Analysis

One strong assumption of conducting the PSM is that there remains no unobserved confounding. It is impossible to prove that no unobserved confounding exists, but through a sensitivity analysis, we can measure if the results are sensitive to hidden bias. There are several sensitivity analysis tests that can be used to check for hidden bias. For binary outcomes, a McNemar's exact test of sensitivity or the Mantel–Haenszel developed by Rosenbaum (2002) can be used. The McNemar's test compares the number of differing pairs in which appropriately CF children had positive outcomes (i.e., not stunted) against non-appropriately fed children who had a negative outcome (stunted) (Hernán & Robins, 2020; Rosenbaum, 2005). On the other hand, the Mantel–Haenszel (MH) non-parametric test statistics can be used to assess the sensitivity of the exposure effect to unmeasured confounder bias. The MH non-parametric test compares the successful number of individuals in the treatment group with the same expected number, given that the treatment effect is zero (Aakvik, 2001).

For a sensitivity analysis, a range of values for the odds ratio of two matched children is considered, which is denoted as *gamma* in the literature (Rosenbaum, 2005; Aakvik, 2001). The gamma can take a range of values with the common one being from 1 to a maximum value of 2 with several increment values (i.e., 0.2). If there were no unmeasured confounding, this maximum odds ratio would be 1; higher values of the maximum odds ratio correspond to more unmeasured confounding (Keele, 2010).

Let Q^+_{MH} be the test statistic, given that we have overestimated the treatment effect, and Q^-_{MH}, the case where we have underestimated the treatment effect. The two bounds for the MH are then given by

$$Q^+_{MH} = \frac{|Y_1 - \sum_{s=1}^{S} \widetilde{E}^+_s| - 0.5}{\sqrt{\sum_{s=1}^{S} Var(\widetilde{E}^+_s)}} \tag{16}$$

or

$$Q_{MH}^{-} = \frac{|Y_1 - \sum_{s=1}^{S} \widetilde{E}_s^{-}| - 0.5}{\sqrt{\sum_{s=1}^{S} Var(\widetilde{E}_s^{-})}}. \tag{17}$$

Mantel–Haenszel (MH) tests calculate the bounds to check sensitivity of the ATE weight results (Aakvik, 2001).

The McNemar's test assesses the null hypothesis of no appropriate complementary effect on child growth for different values of unobserved heterogeneity Rosenbaum (2005) by computing an upper and lower bounds using $p^{+} = \frac{\Gamma}{1+\Gamma}$ and $p^{-} = \frac{1}{(1+\Gamma)}$.

The McNemar's non-parametric test compares the number of discordant pairs in which appropriately CF children had improved growth against non-appropriately fed children who did not have improved growth (Rosenbaum, 2002). The statistic is repeated for different values of Γ to find the value of Γ at which the upper bound p-values become non-significant ($p > 0.05$). The upper bound p-value is given as

$$\sum_{a}^{N} (p^{+})^{a} (1 - p^{+})^{N-a}.$$

N is the total number of discordant pairs and a is the discordant pairs in which children who were not appropriately CF were wasted or stinted or underweight and those who were not appropriately CF were not.

The sensitivity analysis aims to assess how the inference about the appropriate complementary effect will be altered by changing the values of Γ that measures unobserved covariates (Rosenbaum, 2005).

In this chapter, the level of gamma (Γ), a range of possible values attributable to unobserved heterogeneity, was set from 1 to 2 with an increment of 0.1. A value of gamma close to 1 and significant indicates sensitivity to unobserved heterogeneity (Rosenbaum, 2005). Several packages are available to conduct a sensitivity test on the causal effects for binary outcomes in R Rosenbaum and Small (2017) and Stata (Becker & Caliendo, 2007). We used the *mhbounds* package in Stata to obtain the values of the upper and lower bounds on the estimates for wasting, stunting, and underweight (Keele, 2010).

3 Illustrative Examples Using Nutritional Outcome Data in Children

Population- and household-based surveys such as the Demographic and Health surveys are nationally representative household surveys that are widely used and provide information on a wide range of indicators in the areas of population,

maternal and child health, and nutrition (Manda et al., 2014). However, a few studies have assessed causal association using this health survey data (Twabi et al., 2020).

3.1 Data

We extracted child data from the 2015–16 Malawian Demographic and Health Survey (MDHS) and the 2018 Zambian Demographic and Health Survey (ZDHS). These are nationally representative household surveys that provided up-to-date information on current HIV trends, maternal health, HIV and AIDS, and child health and nutritional status of children under age 5 (via weight and height measurements) at national level for both rural and urban areas of the country. Both surveys followed a stratified two-stage sample design. Parents of infants included in the surveys had signed an informed consent. Details of the sampling design for the MDHS and ZDHS can be obtained from the 2015–16 MDHS and 2018 ZDHS reports (Government of Malawi and ICF, 2017; Zambia Statistics Agency and ICF, 2019).

For the effect of exclusive breastfeeding, we analysed separately data of 1978 and 1182 infants less than 6 months of age from the 2015–16 Malawian and 2018 Zambian surveys, respectively. To estimate the effect of appropriate complementary feeding (ACF) on the child nutritional outcomes, analysis was done on data of 4722 and 2879 children aged 6 to 23 months from the 2015–16 MDHS data and the 2018 ZDHS, respectively. In Malawi, HIV test results were not made available to respondents, while in Zambia, HIV testing was performed in households of which respondents chose to be informed of their HIV test result (Government of Malawi and ICF, 2017; Zambia Statistics Agency and ICF, 2019).

3.1.1 Causal Pathway Framework

Figure 1 shows the causal pathway of the effect of a child's feeding practice (EBF or ACF) on child nutritional outcomes. Several variables were identified from the ZDHS and MDHS as potential confounders on the effect of exclusive breastfeeding on child nutritional outcomes, and these include maternal age, maternal HIV, wealth status, place of residence, maternal education, sex of a child, size of a child at birth, and age of a child, mothers employment, duration of breastfeeding, counselling on breastfeeding, birth weight, and having diarrhoea (Ali et al., 2017; Ayisi & Wakoli, 2014; Woldeamanuel & Tesfaye, 2019; Chekol et al., 2017). For appropriate complementary feeding, the potential confounders identified included place of residence, wealth status, region of stay, antenatal care visits, access to media, maternal age, maternal HIV, counselling on feeding, and child's age (Kassa et al., 2016; Kamenju et al., 2017; Perkins et al., 2018; Walters et al., 2019; Disha et al., 2012). All the confounders were compared between the treated (EBF or ACF) and control (no EBF or no ACF) groups using a *chi-square test* for categorical variables and a *t-test* for continuous variables. Mothers HIV infection

APC—appropriate complementary feeding, EBF—exclusive breastfeeding, Und—underweight, Was—wasting, Stu—stunting, Mage—maternal age, CSex—sex of a child, BW—birth weight, Medu—mothers' education, Res—place of residence, ANC—antenatal care visits

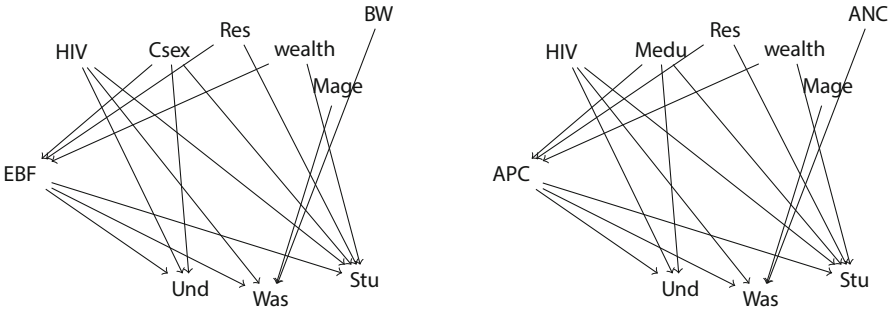


Fig. 1 **Left:** Causal pathway for exclusive breastfeeding and child nutritional outcomes (0–6 months). **Right:** Causal pathway for appropriate complementary feeding and child nutritional outcomes (6–23 months)

is known to have an effect on child growth only and can be defined as a potential confounder. As shown in Fig. 1, we postulate a path from HIV status to the outcomes only. However, for the ZDHS data, mothers were informed of their HIV status; hence, the decision to either exclusive breastfeed a child or practice appropriate complementary feeding may be influenced by the mothers' HIV status. Therefore, we assessed the influence of maternal HIV on the causal associations for the ZDHS data. The causal paths are shown from feeding to the nutritional outcomes wasting, stunting, and underweight. The paths indicating adjusting for confounders are arrows drawing from confounders into the variable EBF or ACF and the outcomes (stunting, wasting, or underweight). The bias pathway is shown from the arrows that show a relationship between HIV infection and stunting, wasting, or underweight.

3.1.2 Exclusive Breastfeeding and Complementary Feeding Indicators

The exposure variables were exclusive breastfeeding (EBF) and appropriate complementary feeding (ACF) categorised as (0=exclusively breastfed, 1=not exclusively breastfed) and (0=not appropriate, 1=appropriate), respectively. The Demographic and Health Surveys (DHS) have information on a one day (24 h) infant diet recall method, and this was used for assessing exclusive breastfeeding. Exclusive breastfeeding was calculated if an infant was fed only breast milk (with the exception of ordered medicines and vitamins by health professionals) one day (24 h) before the survey was conducted.

Complementary feeding practices were measured using the key indicators recommended by the WHO/UNICEF in 2008 that include introduction of solid, semi-solid, or soft foods, minimum dietary diversity, minimum meal frequency

and minimum acceptable diet calculated for the age ranges 6–11, 12–17, and 18–23 months of age, and based on a 24-h recall of the child's dietary intake. These indicators include:

(i) Introduction to complementary foods = 1 if a child aged 6–23 months was complementary fed (solid, semi-solid, or soft) and 0 otherwise (WHO et al., 2010; Kassa et al., 2016).
(ii) Minimum dietary diversity (MDD) = 1 if a child received foods from four or more food groups during the previous day and 0 otherwise. This refers to the child receiving the following food groups: grains, roots, and tubers; legumes and nuts; dairy products (milk, yoghurt, and cheese); flesh foods (meat, fish, poultry, and liver/organ meats); eggs; vitamin A-rich fruits and vegetables; and other fruits and vegetables (WHO et al., 2010; Kassa et al., 2016).
(iii) Minimum meal frequency (MMF)=1 if a breastfeeding and non-breastfeeding child aged 6–23 months received complementary foods the minimum number of times or more (minimum was defined as: two times for breastfed infants 6–8 months; three times for breastfed children 9–23 months; and four times for non-breastfed children 6–23 months) in the previous day (WHO et al., 2010; Kassa et al., 2016), 0 otherwise.
(iv) Minimum acceptable diet (MAD) = 1 if a child was fed a minimum dietary diversity and minimum meal frequency during the day or night preceding the survey (WHO et al., 2010; Government of Malawi and ICF, 2017) and 0 otherwise. Minimum acceptable diet was calculated as composite indicator from the following:

    a. Breastfed children—minimum dietary diversity and minimum meal frequency as above
    b. Non-breastfed children—minimum dietary diversity but excluding the dairy products category (4 out of 6 groups) and minimum meal frequency and 2 or more milk feeds (Government of Malawi and ICF, 2017)

Appropriate complementary feeding was quantified by considering core WHO infant and young child feeding indicators from the DHS data. If a child met these core indicators, introduction of complementary feeding, minimum dietary diversity, and minimum meal frequency, then the child was classified as having received appropriate complementary foods.

3.2 Results

In this chapter, the analysis focused on two interventions, namely: exclusive breastfeeding and appropriate complementary feeding. In Malawi, the prevalence of exclusive breastfeeding among children aged 0–6 months was 52.4% (95% CI: (50.2%, 55.0%)), and in Zambia, it was 64.4% (95% CI: (58.1%,68%)). A majority of children aged 6 to 23 months in Malawi (83.2%(82.1%,84.3%)) and Zambia

96.5%(95.7%,97.2%) were on complementary feeding. The prevalence of children who received appropriate complementary feeding in Malawi was 7.8% (7.0%, 8.5%) and 11.4% (10.0%, 12.9%) in Zambia.

Before application of the propensity score methods, we explore the distribution of the confounders across the intervention groups as presented in Tables 1 and 2. We observe that in Malawi among the 1012 children who were exclusively breastfed, 25 (7.8%) were born to HIV-infected mothers, 835 (84.1%) resided in the rural part of Malawi, and 437 (45%) were from poor households. In Zambia, among the 762 children who were exclusively breastfed, 73 (10.1%) were born to HIV-infected mothers, 370 (48.6%) were female, 51 (6.7%) had reported having diarrhoea two weeks before the survey, and 598 (78.5%) were aged 0–3 months, see Table 1. We observe that mothers' education, history of diarrhoea, history of vitamin A uptake, mothers' age, birth weight, and child's age were significantly different between exclusively breastfed and non-exclusively breastfed children in Malawi. In Zambia, sex of a child, place of residence, history of diarrhoea, vitamin A uptake, and child's age were significantly different between the EBF groups.

Table 1 further presents absolute standardised differences (ASD) for a confounder across the intervention groups. An ASD value of 0.1 (10%) was used as a decision criterion and denoted meaningful balance in the measured covariates (Normand et al., 2001; Twabi et al., 2020). A closer look at the ASDs in the table shows that some of the confounders had an ASD > 0.1, implying imbalance in the EBF groups among the children.

Table 2 presents distribution of confounders among appropriate complementary fed children. Among the 368 children who were on ACF in Malawi, 44.3% were females, while among the 4354 children not on ACF, 43.3% were female. In Zambia, among the 286 children on ACF, 164 (51.9%) resided in the rural parts of Zambia, while 1842 (67.3%) of the 2593 who were not on ACF were from the rural part of Zambia. In Malawi, maternal HIV, place of residence, sex of a child, mothers' age, wealth index, and mothers' education were significantly different between children on ACF and no ACF. For the ZDHS data, place of residence, wealth index, mothers' education, and age of a child were significantly different between the ACF groups. The ASD differences for these confounders are also greater than 0.1, affirming to the imbalance of the confounders between the two groups.

3.2.1 The Effect of the Nutritional Interventions on Child Nutritional Outcomes Before PS Application

Table 3 presents the effects of exclusive breastfeeding and appropriate complementary feeding on the child nutritional outcomes before applying the propensity score matching and PS weighting. There was a positive association between exclusive breastfeeding and appropriate complementary feeding on the child growth indicators; however, the association was not statistically significant. In Malawi, children who were on exclusive breastfeeding were less likely to be wasted and underweight, while for Zambia, they were less likely to be stunted, wasted, and

Table 1 Distribution of confounders between EBF groups for children aged 0–6 months before matching

	2015–16 MDHS				2018 ZDHS			
	EBF (n=1012)	No EBF (n=966)			EBF (n=762)	No EBF (n=420)		
Characteristic	n(%)	n(%)	*p*	ASD	n(%)	n(%)	*p*	ASD
Mother HIV status								
HIV negative	295(92.2)	284(92.9)		0.008	656(89.9%)	365(88.8)		0.02
HIV positive	25(7.8)	25(7.1)	0.897	0.006	73(10.1%)	46(11.2)	0.6068	0.05
Sex of child								
Male	484(47.2)	494(50.7)		0.07	392(51.4)	182(43.4)		0.16
Female	528(52.8)	472(49.3)	0.141	0.07	370(48.6)	238(56.6)	0.0391	0.16
Place of residence								
Urban	177(15.9)	164(13.4)		0.014	248(32.5)	151(36)		0.07
Rural	835(84.1)	835(86.6)	0.769	0.11	514(67.5)	269(64)	0.029	0.07
Mothers' education								
None	125(13.1)	113(11.8)		0.02	80(10.5)	39(9.2)		0.04
Primary	633(62.6)	658(69.4)		0.12	386(50.6)	196(46.5)		0.08

(continued)

Table 1 (continued)

Characteristic	2015–16 MDHS				2018 ZDHS			
	EBF (n=1012)	No EBF (n=966)			EBF (n=762)	No EBF (n=420)		
	EBF (n=1012)	No EBF (n=966)			EBF (n=762)	No EBF (n=420)		
	n(%)	n(%)	*p*	ASD	n(%)	n(%)	*p*	ASD
Post-primary	254(24.3)	195(18.9)	0.02	0.12	296(38.9)	186(44.3)	0.286	0.11
Had diarrhoea								
No	928(90.9)	758(77.9)		0.38	711(93.3)	333(79.2)		0.41
Yes, 2 weeks ago	84(9.1)	208 (22.1)	<0.001	0.38	51(6.7)	88(20.8)	<0.001	0.41
Had vitamin A								
No	703(70.5)	543(59.6)		0.27	697(91.5)	340(80.9)		0.31
Yes	309(29.5)	423(40.4)	0.001	0.28	65(8.5)	80(19.1)	<0.001	0.31
Wealth								
Poor	437(45)	449(49.7)		0.07	376(49.4)	205(48.9)		0.01
Middle	186(18.9)	190(19.1)		0.03	147(19.3)	81(19.3)		0.00
Rich	389(36.1)	327(31.2)	0.105	0.10	239(31.3)	134(31.8)	0.647	0.01
Mothers' age								
15–24	480(48.3)	509(53.0)		0.10	342(44.9)	192(45.7)		0.02
25–34	372(36.9)	356(36.7)		0.002	295(38.7)	178(42.3)		0.07
35 above	160(14.8)	101(10.3)	0.001	0.16	125(16.4)	50(11.9)	0.255	0.13
Birth weight								
Normal	926(91.2)	861(88.7)		0.08	709(93.1)	394(93.7)		0.03
Low birth	86(8.8)	105(11.3)	0.011	0.08	53(6.9)	27(6.3)	0.729	0.02
Child's age								
0–3 months	795(79.2)	314(31.8)		1.05	598(78.5)	88(20.8)		1.4
4–6 months	217(20.8)	652(68.2)	<0.001	1.05	164(21.5)	333(79.2)	<0.001	1.4

n—Number of observations
%—Percent
ACF—Appropriate complementary feeding
p—p-value
ASD—Absolute standardised difference

Table 2 Distribution of confounders between appropriate complementary groups for children aged 6–23 months before matching

	2015-16 MDHS				2018 ZDHS			
	ACF (n=368)	No ACF (n=4354)			ACF(n=286)	No ACF (n=2593)		
Characteristic	n(%)	n(%)	*p*	ASD	n(%)	n(%)	*p*	ASD
Maternal HIV								
HIV-uninfected	161(43.8%)	161(43.8%)		0.05	255(90.4%)	2282 (87.0%)		0.06
HIV-infected	207(56.2%)	2560(58.8%)	0.038	0.3	22 (6.9%)	242 (10.2%)	0.326	0.04
Sex of a child								
Male	205(55.7%)	2467(56.7%)		0.02	134 (48.8%)	1288 (50.5%)		0.05
Female	163(44.3%)	1887(43.3%)	0.05	0.02	152 (52.2%)	1305 (49.5%)	0.3216	0.06
Residence								
Urban	136(36.9%)	1028(23.6%)		0.29	122(48.1%)	751(32.7%)		0.29
Rural	232(63.1%)	3326(76.4%)	<0.001	0.29	164(51.9%)	1842(67.3)	<0.001	0.29
Mothers' age								
15–24	156(42.4%)	1940(44.6%)		0.04	102 (33.7%)	1099 (42.3%)		0.14
25–34	161(43.8%)	1726(40.6%)		0.08	123 (42.7%)	1006 (38.7%)		0.09
35 above	51(13.9%)	926(15.8%)	0.005	0.20	61 (23.6%)	488 (19%)	0.064	0.06

(continued)

Table 2 (continued)

Characteristic	2015-16 MDHS				2018 ZDHS			
	ACF (n=368)	No ACF (n=4354)			ACF(n=286)	No ACF (n=2593)		
	n(%)	n(%)	*p*	ASD	n(%)	n(%)	*p*	ASD
Wealth								
Poor	101(27.5%)	2075(47.7%)		0.43	104 (34.8%)	1401 (49.7%)		0.36
Middle	61(16.6%)	827(18.9%)		0.06	59 (15.9%)	528(18.8%)		0.01
Rich	206(55.9%)	1452(33.4%)	<0.001	0.34	123 (49.3%)	664 (31.5%)	<0.001	0.37
Education								
None	45(12.2)	811(18.6)		0.18	15 (4.8)	272 (9.9)		0.20
Primary	194(52.7)	2745(63.1)		0.21	126 (42.4)	1373 (52.0)		0.18
Secondary & Post-sec	129(35.1)	798(18.3)	<0.001	0.39	145 (52.8)	948 (38.1)	<0.001	0.29
Antenatal visits								
0–2 times	34 (10.4)	608 (14.5)		0.15	20(7.3)	237 (9.5)		0.08
3–4 times	133 (36.3)	1605 (36.5)		0.01	171 (59.4)	1497 (60.3)		0.04
More than 5	197 (53.3)	2136 (49.0)	0.215	0.09	95 (33.3)	785 (30.3)	0.524	0.06
Age of a child								
6–11 months	114 (31)	1485(34.1)		0.07	100 (36.7)	851 (32.3)		0.05
12–17 months	131 (35.6)	1518 (34.9)		0.02	134 (45.7)	830 (32.4)		0.31
18–23 months	123 (33.5)	1351 (31.0)	0.231	0.05	52 (17.5)	912 (35.2)	<0.001	0.39
6–23 months	368 (100)	4354 (100)			286 (100)	2593 (100)		

n—Number of observations
%—Percent
ACF—Appropriate complementary feeding
p—p-value
ASD—Absolute standardised difference

underweight, see Table 3. Appropriately complementary fed children were less likely to be wasted in Malawi and were less likely to be wasted, stunted, and underweight in Zambia, see Table 3.

3.2.2 Estimating the Propensity Score

All covariates that were identified as potential confounders were included in a logistic model to estimate the probability of being exclusively breastfed (propensity scores for EBF) and the probability of being on appropriate complementary feeding (propensity scores for ACF). A forward model selection was applied to select potential interactions. The Hosmer and Lemeshow goodness-of-fit test was performed to check if the model fits the data well. The p-values for the Hosmer and Lemeshow test on exclusive breastfeeding were 0.4632 for the MDHS data and 0.4632 for the ZDHS data, and for appropriate complementary feeding, the p-values were 0.5881 for the MDHS data and 0.1228 for the ZDHS data, indicating the treatment models fit the data well.

3.2.3 PS Matching

Exclusively breastfed children were matched to those who were not exclusively breastfed and children who were appropriately complementary fed to those who were not appropriately complementary fed. The differences on the matched sample for both datasets were explored after matching using the McNemar's test, and the results are presented in Tables 4 and 5. For appropriate complementary feeding, a sample of 368 matched pairs was created for the MDHS data and a sample of 283 matched pairs was created for the ZDHS data. For exclusive breastfeeding, a sample of 390 and 537 matched pairs was created for Zambia and Malawi, respectively. The pairs were matched using nearest neighbour matching within a 0.01 calliper. Significant differences between children who were EBF and those who were not EBF were balanced for some covariates as seen in the absolute standardised differences (ASD) listed in Table 4. For appropriate complementary feeding, the propensity score matching succeeded in balancing the differences in all the confounders, as seen from the ASD and p-values obtained, between ACF and non-ACF children as presented in Tables 5.

The average treatment effect was then calculated from the matched sample using a conditional logistic regression for binary data and using a paired t-test for the continuous outcomes. In Zambia, children who were exclusively breastfed were less likely to be stunted (OR=0.75, (95% CI: 0.54,1.04)) and underweight (OR=0.84, (95% CI: 0.53, 1.33)) compared to children who were not exclusively breastfed. However, the positive effect was not statistically significant. In Malawi, exclusive breastfeeding had a positive effect on underweight (OR=0.77, (95% CI: 0.39, 1.53)) and wasting (OR=0.69, (95% CI: 0.28,1.71)). However, the positive effect was not statistically significant, see Table 6.

Table 3 The adjusted effect of nutritional interventions by a child's age group before applying PS techniques

(a) Exclusive breastfeeding: 0–6 months

	2015–16 MDHS			2018 ZDHS		
	Stunted	Wasted	Underweight	Stunted	Wasted	Underweight
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
No EBF	1.00	1.00	1.00	1.00	1.00	1.00
EBF	1.40 (0.87, 2.27)	0.80 (0.29, 2.26)	1.36 (0.59, 3.16)	1.10 (0.76, 1.59)	0.89 (0.37, 2.13)	0.93 (0.55, 1.54)

(b) Appropriate complementary feeding: 6–23 months

	2015–16 MDHS			2018 ZDHS		
	Stunted	Wasted	Underweight	Stunted	Wasted	Underweight
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Not appropriate	1.00	1.00	1.00	1.00	1.00	1.00
Appropriate	0.78 (0.46, 1.35)	0.71 (0.21, 1.05)	0.47 (0.21, 1.05)	0.92 (0.70, 1.22)	0.83 (0.45, 1.54)	0.68 (0.44, 1.06)

Table 4 Distribution of confounders between EBF groups for children aged 0–6 months after PS matching

	2015–16 MDHS				2018 ZDHS			
	EBF (n=537)	No EBF (n=537)			EBF (n=390)	No EBF (n=390)		
Characteristic	n(%)	n(%)	*P*	ASD	n(%)	n(%)	*P*	ASD
Mothers HIV status								
HIV negative	492 (91.6)	508 (94.6)			357 (91.5)	344 (88.2)		0.023
HIV positive	45 (8.4)	29 (5.4)	0.054	0.11	33 (8.5)	46 (11.8)	0.123	0.026
Sex of a child								
Male	254 (47.3)	261 (48.6)			201 (51.5)	182 (46.7)		0.09
Female	283 (52.7)	276 (51.4)	0.669	0.03	189 (48.5)	208 (53.3)	0.174	0.09
Place of residence								
Urban	116 (21.6)	99 (18.4)	0.07	0.07	117 (30.0)	109 (27.9)		0.06
Rural	421 (78.4)	438 (81.6)	0.195	0.08	273 (70.0)	281 (72.1)	0.528	0.05
Mothers' education								
None	75 (13.9)	61 (11.4)			41 (10.5)	36 (9.2)		
Primary	311 (57.9)	375 (69.8)		0.24	171 (43.9)	194 (49.7)		0.12
Secondary/Post-secondary	151 (28.2)	101 (18.8)	<0.001	0.22	178 (45.6)	160 (41.1)	0.255	0.09

(continued)

Table 4 (continued)

	2015–16 MDHS				2018 ZDHS			
	EBF (n=537)	No EBF (n=537)			EBF (n=390)	No EBF (n=390)		
Characteristic	n(%)	n(%)	*P*	ASD	n(%)	n(%)	*P*	ASD
Had diarrhoea								
No	455 (84.7)	464 (86.4)			343 (87.9)	346 (88.7)		0.00
Yes, last 2 weeks	82 (15.3)	73 (13.6)	0.435	0.05	47 (12.1)	44 (11.3)	0.738	0.024
Had vitamin A								
No	334 (62.2)	326 (60.7)			332 (85.4)	328 (84.5)		
Yes, last 2 weeks	203 (37.8)	211 (39.3)	0.616	0.03	57 (14.7)	60 (15.5)	0.752	0.02
Wealth								
Poor	211 (39.3)	252 (46.9)			152 (38.9)	208 (53.3)		
Medium	97 (18.1)	103 (19.2)		0.03	94 (24.1)	76 (19.5)		0.11
Rich	229 (22.6)	182 (33.9)	0.01	0.18	144 (36.9)	106 (27.2)	<0.001	0.21
Mothers' age								
15–24 years	294 (54.8)	251 (46.7)			166 (42.6)	182 (46.7)		
25–34 years	156 (29.1)	218 (40.6)		0.24	144 (36.9)	159 (40.8)		0.08
35–49 years	87 (16.2)	68 (12.7)	<0.001	0.1	80 (20.5)	49 (12.6)	0.012	0.21
Child's age								
0–3 months	260 (48.4)	251 (46.7)			256 (65.4)	95 (24.4)		
4–6 months	277 (51.6)	286 (53.3)	0.582	0.03	135 (34.6)	295 (75.6)	<0.001	0.9

n—Number of observations
%—Percent
ACF—Appropriate complementary feeding
p—p-value
ASD—Absolute standardised difference

Table 5 Distribution of confounders between ACF groups for children aged 6-23 months after PS matching

	2015–16 MDHS				2018 ZDHS			
	ACF (n=368)	No ACF (n=368)			ACF (n=283)	No ACF (n=283)		
Characteristic	n(%)	n(%)	*P*	ASD	n(%)	n(%)	*P*	ASD
Maternal HIV								
HIV-uninfected	161(43.8)	162 (44.0)			250 (88.3)	248 (87.6)		
HIV-infected	207(56.2)	206 (56.0)	0.534	0.05	23 (8.2)	32 (11.3)	0.534	0.10
Sex of a child								
Male	205(55.7)	213 (57.9)			125 (44.2)	123 (43.5)		
Female	163(44.3)	155 (42.1)	0.2864	0.06	158 (55.8)	160 (56.5)	0.865	0.01
Residence								
Urban	136(36.9)	137 (37.2)			121 (42.8)	121 (42.8)		
Rural	232(63.1)	231 (62.8)	0.6617	0.05	162 (57.2)	162 (57.2)	0.46	0.00

(continued)

Table 5 (continued)

	2015–16 MDHS				2018 ZDHS			
	ACF (n=368)	No ACF (n=368)			ACF (n=283)	No ACF (n=283)		
Characteristic	n(%)	n(%)	*P*	ASD	n(%)	n(%)	*P*	ASD
Mothers' age								
15–24	156(42.4)	167 (45.4)			97 (34.2)	96 (33.9)		
25–34	161(43.8)	149 (40.5)		0.03	127 (44.9)	118 (41.7)		0.06
35 above	51(13.9)	52 (14.1)	0.9728	0.02	59 (20.9)	69 (24.4)	0.572	0.08
Wealth								
Poor	101(27.5)	100 (27.2)			90 (31.8)	97 (34.3)		
Middle	61(16.6)	66 (17.9)		0.02	52 (18.4)	49 (17.3)		0.03
Rich	206(55.9)	202 (54.9)	0.3369	0.01	141 (49.8)	137 (48.4)	0.815	0.03
Education								
None	45(12.2)	39 (10.6)			11 (3.9)	18 (6.4)		
Primary	194(52.7)	204 (55.4)		0.07	124 (43.8)	125 (44.2)		0.007
Secondary & Post-secondary	129(35.1)	125 (34)	0.6123	0.08	148 (52.3)	140 (49.5)	0.384	0.06
Age of child								
6–11 months	114 (31)	118 (32.1)			96 (33.9)	114 (40.3)		
12–17 months	131 (35.6)	130 (35.3)		0.05	133 (47.0)	101 (35.7)		0.04
18–23 months	123 (33.4)	120 (32.6)	0.1624	0.00	54 (19.1)	68 (24.0)	0.05	0.12
6–23 months	368 (100)	368 (100)			283 (100)	283 (100)		

n—Number of observations
%—Percent
ACF—Appropriate complementary feeding
p—p-value
ASD—Absolute standardised difference

For the continuous outcomes, in Malawi, children who were exclusively breastfed had an increase in weight-for-height *z-scores* (Coef=3.59, (95% CI: 0.42 6.77)), weight-for-age *z-scores* (Coef=0.38, (95% CI: −1.55, 2.31)), and height-for-age *z-scores* (Coef=0.62, (95% CI: −3.13, 1.89)). However, the effect of EBF was significant only for weight-for-height *z-scores*. In Zambia, there was a positive effect of EBF on all the *z-scores* (weight-for-height *z-scores* (Coef=2.29, (95% CI: −0.47, 5.05)), weight-for-age *z-scores* (Coef=0.32, (95% CI: 0.08, 0.52)), and height-for-age *z-scores* (Coef=0.15, (95% CI: −0.05, 0.35))); however, the effect was significant only for weight-for-age *z-scores*, see Table 7.

3.2.4 PS Weighting

The inverse probability treatment weights (IPTW) were calculated by taking the inverse of the probability of being exclusively breastfed or probability of being ACF conditional on the observed potential confounders. The IPTWs ranged between 0.4 and 7.5 for the Zambia data and from 0.5 to 18.8 for the Malawi data. Using the weights, the absolute standardised difference was calculated on the weighted sample for each covariate, and the results are presented in Table 8. The PS weighting achieved confounder balance between EBF groups for both Malawi and Zambia as shown in Table 8a. The ASDs for all covariates in Malawi were less than 0.1. We observe a similar pattern for appropriate complementary feeding for both Malawi and Zambia, see Table 8b.

The average treatment effect was then estimated by weighting the mean z-score (probability of being stunted) in the exclusively breastfed group (ACF) and subtracting the mean z-score (probability of being stunted) in the non-exclusively breastfed group (No ACF). In addition, we used robust standard error estimation to account for the weights. For the binary data, in Malawi, EBF had a positive effect on wasting (OR=0.85 (95% CI=0.24, 1.67)) and underweight (OR=0.54 (95% CI=0.24, 1.19)). However, the positive effect was not significant. In Zambia, EBF had a positive effect on underweight (OR=0.75 (95% CI=0.44, 1.27)) and stunting (OR=0.84 (95% CI=0.58, 1.2)); however, the effect was not significant, Table 6.

Table 7 presents the EBF effect on the continuous outcomes. In Malawi, children who were exclusively breastfed had a slight increase in their weight-for-height *z-score* (Coef=3.3, (95% CI:0.33, 6.3)), weight-for-age *z-score* (Coef=0.93, (95% CI: −1.04, 2.9)), and height-for-age *z-scores* (Coef=0.11, (95% CI: −2.75, 2.98)). The effect of EBF was significant only for weight-for-height z-scores. In Zambia, there was an increase in the weight-for-age *z-scores* (Coef=0.41, (95% CI: 0.08,0.73)), weight-for-height *z-scores* (Coef=1.48, (95% CI: −0.80,3.76)), and height-for-age z-scores (Coef=0.18, (95% CI: −0.05,0.35)) among children who were exclusively breastfed. However, the significant effect of EBF on the z-scores was observed for weight-for-age *z-scores*.

Table 6 The effect of exclusive breastfeeding on binary child nutritional outcomes among children aged 0–6 months after applying PS techniques

	2015–16 MDHS			2018 ZDHS		
	Wasting	Underweight	Stunting	Wasting	Underweight	Stunting
	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)
PS matching	0.69(0.28,1.71)	0.77(0.39,1.53)	1.10(0.71,1.70)	1.33(0.71,2.50)	0.84(0.53,1.33)	0.75(0.54,1.04)
IPW	0.85(0.24,1.67)	0.54(0.24,1.19)	1.2(0.70,1.71)	1.13(0.55, 2.30)	0.75(0.44, 1.27)	0.84(0.58, 1.20)

Table 7 The effect of exclusive breastfeeding on continuous child nutritional outcomes among children aged 0–6 months after applying PS techniques

	2015–16 MDHS			2018 ZDHS		
	WHZ	WAZ	HAZ	WHZ	WAZ	HAZ
	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)
PS matching	3.59(0.42, 6.77)	0.38(−1.55,2.31)	0.62(−3.13,1.89)	2.29(−0.47,5.05)	0.32(0.08,0.55)	0.15(−0.05,0.35)
IPW	3.30(0.33,6.3)	0.93(−1.04,2.9)	0.11(−2.75,2.98)	1.48(−0.80,3.76)	0.41(0.08,0.73)	0.18(−0.05,0.35)

WHZ—Weight-for-height z-score
WAZ—Weight-for-age z-score
HAZ—Height-for-age z-score
Coef—Coefficient
CI—Confidence Interval

Table 8 Absolute standardised differences for the confounders across the interventions groups after PS weighting

(a) Exclusive breastfeeding (EBF vs. No EBF):0–6 months			(b) Appropriate complementary feeding (ACF vs. no ACF): 6–23 months		
	2015–16 MDHS	2018 ZDHS		2015–16 MDHS	2018 ZDHS
Confounder	ASD	ASD	Confounder	ASD	ASD
Sex of a child			Maternal HIV		
Male	0.07	0.008	HIV negative	0.003	0.013
Female	0.07	0.008	HIV positive	0.003	0.013
Maternal HIV			Sex of a child		
HIV negative	0.015	0.001	Male	0.04	0.09
HIV positive	0.015	0.001	Female	0.04	0.09
Mothers' education			Mothers' education		
None	0.04	0.19	None	0.04	0.19
Secondary	0.04	0.19	Secondary	0.04	0.19
Post-secondary			Post-secondary		
Child's age			Child's age		
0–3 months	0.019	0.00	6–11 months	0.02	0.04
4–6 months	0.019	0.00	12–17 months	0.016	0.05
			18–23 months	0.039	0.01
Residence			Residence		
Urban	0.034	0.026	Urban	0.01	0.05
Rural	0.034	0.026	Rural	0.01	0.15
Had diarrhoea			Wealth		
No	0.021	0.001	Poor	0.07	0.05
Yes	0.021	0.001	Medium	0.04	0.01
			Rich	0.05	0.01
Had vitamin A			Mothers' age		
No	0.006	0.012	15–24 years	0.07	0.07
Yes	0.0006	0.012	25–34 years	0.04	0.06
			35–49 years	0.05	0.005
Wealth			Antenatal visits		
Poor	0.01	0.09	0–2 times	0.07	0.07
Medium			3–4 times	0.04	0.01
Rich	0.04	0.19	More than 5 times	0.003	0.03
Mothers' age					
15–24 years	0.018	0.017			
25–34 years	0.005	0.024			
35–49 years	0.019	0.003			
Birth weight					
Normal	0.03	0.012			
Low birth weight	0.03	0.012			

ASD—Absolute standardised difference

Table 9 The effect of maternal HIV on the causal association of EBF on nutritional outcomes among children aged 0–6 months in Zambia

	2018 ZDHS		
	Wasting	Stunting	Underweight
	OR(95% CI)	OR(95% CI)	OR(95% CI)
Unadjusted	0.6(0.1,4.6)	2.7 (0.8, 8.9)	1.7(0.1,31.5)
Matching	0.8(0.1,5.9)	3.5 (0.7, 18.2)	1.1(0.1,8.9)
Weighting	0.6(0.1,4.8)	4.4 (1.1, 18.8)∗	0.5(0.03,9.8)

$*p < 0.05$
CI—Confidence Interval

3.2.5 Effect of Mothers' HIV Status on the Causal Effect of Exclusive breastfeeding for Zambian Children

In Zambia, mothers who were tested for HIV were informed of their test result (Zambia Statistics Agency and ICF, 2019). This may have influenced the decision to practice exclusive breastfeeding. We assessed the modifying effect of mothers' HIV status on the causal association between exclusive breastfeeding and the child growth outcomes. The aim was to assess whether the outcomes varied between children born to HIV-infected mothers and those born to HIV-uninfected mothers for the Zambia survey. We tested the interaction between EBF and HIV infection status using a logistic regression model. A non-significant interaction effect implied no moderating effect of HIV exposure on the effect of EBF on the child nutritional outcomes.

Children who were exclusively breastfed and were born to HIV-infected mothers were more likely to be stunted than children on EBF who were born to HIV-uninfected mothers (OR=4.4, 95% CI: 1.1,18.8). However, maternal HIV infection had no significant effect on the causal association between exclusive breastfeeding and wasting and underweight using IPW and PS matching, see Table 9.

3.2.6 Effect of Appropriate Complementary Feeding on Child Growth

Further analysis was done on children aged 6 to 23 months to examine the effect of appropriate complementary feeding on child nutritional outcomes using a conditional logistic regression on the matched data and a weighted logistic regression using PS weights. Tables 10 and 11 present results for the effect of appropriate complementary feeding the child growth indicators (binary and continuous). Appropriate complementary feeding had a positive effect on wasting, stunting, and underweight for both data and balancing methods. For the matched sample, appropriate complementary feeding had a positive effect on stunting (OR=0.7, (95% CI: 0.4,0.95)), wasting, and underweight but was only statistically significant for stunting. In Zambia, using PS matching and IPW, appropriate complementary feeding had a positive effect on underweight (OR=0.64,(95% CI: 0.37,1.1)) and stunting (OR=0.87,(95% CI: 0.6,1.24)). However, a significant effect of ACF on stunting was observed for the IPW (OR=0.9,(95% CI: 0.4,0.95)). The estimates for

the PS matching and IPW were different, with a higher effect observed for the IPW as compared to estimates for PS matching (Table 10).

For the continuous outcomes, using the PS matching and IPW, in Malawi, children who were appropriately complementary fed had an increase in their height-for-age *z-scores* (Coef=0.89,(95% CI: −0.78,2.54)), weight-for-age *z-score* (Coef=0.17,(95% CI: −0.11,0.45)) using PS matching and (Coef=0.89,(95% CI: −0.78,2.54)), (Coef=0.28,(95% CI: −0.26, 0.29)) using PS weighting, respectively. In Zambia, children who were appropriately complementary fed had an increase in their weight-for-height *z-scores* (Coef=0.25,(95% CI: −1.6,2.1)) using PS matching and (Coef=0.6,(95% CI: −0.89,0.16)) PS weighting, respectively (Table 11).

3.2.7 Sensitivity Analysis

Table 12a presents the Mantel–Haenszel bounds for exclusive breastfeeding on the nutritional outcomes for the 2018 ZDHS. For wasting, the critical value of overestimating the causal association among 0–6 months aged Zambian children was somewhere below 2 (p =0.08) or 2.2 (p =0.045). This suggests that the results were robust against unobserved confounder bias.

Table 12b presents the results of the Mantel–Haenszel for the 2015–16 MDHS. For stunting, the observed exposure effect would change due to unobserved confounders at an odds ratio of 1 ($p < 0.001$) or 1.6 ($p = 0.026$). We note that the EBF effect on underweight and stunting is unstable. This suggests that the results are prone to underestimation by unobserved confounders. However, for outcome wasting, the critical value of overestimating the causal association was obtained at an odds ratio of 2.2 ($p < 0.001$) or 2.4 ($p < 0.001$). This implies that the results on wasting were robust against unobserved confounder bias. Table 13 in the appendix presents the sensitivity analysis after matching on children aged 6–23 months. For both the MDHS and ZDHS data, the effect of appropriate complementary feeding on stunting was prone to change due to unobserved confounders as compared to underweight and wasted. Therefore, there is a need to observe caution when interpreting the results for stunting as causal.

4 Discussion

Assessing causal association is an important research objectivity in health research and primarily relies on findings from conduction of a randomised control trial (RCT). However, due to the impracticability of RCTs for some health problems, observational studies are widely used to assess causal association of an exposure on health outcomes. In this chapter, we have described statistical underpinnings of propensity score matching and inverse probability weighting that can be used to reduce confounder bias and ensure the estimation of causal association from observational studies. The descriptions of these methods have been complemented by an

Table 10 The effect of appropriate complementary feeding on the binary child nutritional outcomes among children aged 6–23 months

	2015–16 MDHS			2018 ZDHS		
	Wasting	Underweight	Stunting	Wasting	Underweight	Stunting
	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)
PS matching	0.4 (0.1,1.7)	0.6 (0.2,1.7)	0.7 (0.4,0.95)	1.54(0.62,3.83)	0.64(0.37,1.11)	0.87(0.61,1.24)
IPW	0.65(0.25,1.68)	0.63(0.30,1.31)	0.89(0.68,1.38)	1.2(0.6, 2.3)	0.8(0.4, 1.6)	0.9(0.4, 0.95)

Table 11 The effect of appropriate complementary feeding on the continuous child nutritional outcomes among children aged 6–23 months

	2015–16 MDHS			2018 ZDHS		
	WHZ	WAZ	HAZ	WHZ	WAZ	HAZ
	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)	Coef.(95%CI)
PS matching	0.09(−0.22,0.39)	0.17(−0.11,0.45)	0.89(−0.78,2.54)	0.25(−1.62,2.12)	0.02(−0.16,0.19)	−0.87(-3.02,1.28)
IPW	−0.03(−0.23,0.22)	0.12(−0.08,0.31)	0.28(−0.26,0.29)	0.60(−0.89,0.16)	−0.06(−0.23,0.13)	−0.36(−1.42,2.14)

WHZ—Weight-for-height z-score
WAZ—Weight-for-age z-score
HAZ—Height-for-age z-score

Table 12 Sensitivity analyses for unobserved confounding for EBF groups after matching among children aged 0–6 months

(a) ZDHS 2018				(b) MDHS 2016-15			
Gamma	Underweight Q-MH^-	Wasting Q-MH^+	Stunting Q-MH^-	Gamma	Q-MH^- Under-weight	Wasting Q-MH^+	Stunting Q-MH^-
1	0.56**	0.21	2.1	1	3.95****	0.81**	5.18****
1.2	0.15*	-0.05	1.18	1.2	3.32****	0.26**	3.89****
1.4	-0.19	0.37	0.42	1.4	2.81***	-0.14*	2.84***
1.6	0.04	0.75	0.04	1.6	2.38***	0.25	1.93**
1.8	0.30	1.08	0.62	1.8	2.01**	0.60	1.13
2	0.53	1.38*	1.14	2	1.68**	0.91	0.42
2.2	0.74	1.65**	1.61*	2.2	1.39	1.20	0.07
2.4	0.94	1.9**	2.05**	2.4	1.12	1.46	0.66
2.6	1.12	2.1**	2.45***	2.6	0.88	1.70	8
2.8	1.29	2.36***	2.83***	2.8	0.66	1.93	1.69
3	1.45	2.57***	3.18****	3	0.45	2.14	2.16

Data imputed using multiple imputation
Q-MH^+ or Q-MH^- Mantel–Haenszel statistic for overestimation and underestimation of the ATE
$*p < 0.10$; $**p < 0.05$; $***p < 0.01$; $****p < 0.001$

evaluation of the effect of exclusive breastfeeding and appropriate complementary feeding on child growth using health survey data in Malawi and Zambia. There was confounder balance between exclusively breastfed and non-exclusively breasted children and between appropriately complementary fed and non-appropriately complementary fed children for both surveys after applying confounder balancing methods. We assessed the causal association using the matched and weighted samples.

The effect of exclusive breastfeeding and appropriate complementary feeding on the growth outcomes for both the MDHS and ZDHS differed between the two based bias correction methods. The confidence intervals (CIs) for the PS matching were slightly wider than the CIs for the IPW. Other studies have shown that the IPW has smaller variances as compared to the PS matching, hence producing narrower CIs (Abadie & Imbens, 2006; Austin, 2011b). Before applying the PS matching and IPW, estimates from the adjusted regression for the binary outcomes showed that exclusive breastfeeding had a non-significant negative association on stunting and underweight for the MDHS data, while for the ZDHS, a non-significant negative effect of exclusive breastfeeding on stunting was observed. After applying the bias correction methods, we observe a positive non-significant effect of EBF on wasting and underweight, while for ZDHS, there was a positive effect of EBF on underweight and stunting. Another observation worth noting is that, after applying the bias correction methods, we observe a change (maximum or minimal) in the

Table 13 Sensitivity analysis for unobserved confounding after matching for appropriate complementary feeding among children aged 6–23 months

(a) 2015–16 MDHS				(b) 2018 ZDHS			
Gamma	Underweight Q-MH^+	Wasted Q-MH^+	Stunted Q-MH^-	Gamma	Underweight Q-MH^-	Wasted Q-MH^+	Stunted Q-MH^-
1.0	0.04**	0.09***	0.09****	1	1.29*	0.33	0.35
1.2	0.13*	0.19**	0.53****	1.2	0.69	−0.10	0.37
1.4	0.16*	0.43**	1.06****	1.4	0.18	0.03	1.15
1.6	0.42	0.64*	1.52****	1.6	−0.05	0.34	1.83*
1.8	0.65	0.83	1.93****	1.8	0.34	0.62	2.42**
2.0	0.85	0.99	2.29***	2	0.68	0.86	2.96***
2.2	1.03	1.15	2.63***	2.2	1.00	1.09	3.45****
2.4	1.21	1.29	2.94**	2.4	1.28*	1.30053*	3.89****
2.6	1.37	1.43	3.22*	2.6	1.55*	1.49471*	4.31****
2.8	1.52	1.55	3.49	2.8	1.79**	1.67635**	4.70****
3.0	1.66	1.67	3.74	3	2.03**	1.84727**	5.06****

Data imputed using multiple imputation
Q-MH^+ or Q-MH^- Mantel–Haenszel statistic for overestimation and underestimation of the ATE.
$*p < 0.10$; $**p < 0.05$; $***p < 0.01$; $****p < 0.001$

magnitude of the effects. For instance, for the MDHS, the positive effect of ACF on stunting was shown to have an odds ratio (OR) of 0.47 (53% reduction); however after applying the bias correction methods, the positive effect of ACF had an OR of 0.89 (11% reduction). Thus, we observe the importance of applying the PS matching and IPW that control for confounder bias when estimating the causal effect from observational studies, as the estimates before applying the PS methods are either underestimated or overestimated and while the PS matching and IPW improve the power to detect the effects.

Estimation of the propensity score and its balancing property were done under the assumption that there were no unmeasured confounders. The estimation of the average causal effect differed for the propensity score matching and the IPW. The propensity score matching used the estimated propensity score to obtain a good matched sample, and the ATE was estimated from this sample. This implies that the propensity score was not directly used for analysis under PS matching. However, the IPW directly used the propensity scores in estimating the average treatment effect. Thus, the propensity score matching may be less sensitive to misspecification of the propensity score model unlike the IPW. It is important to note that all the approaches based on propensity scores can only address observed measured confounders.

In this chapter, we estimated the propensity score using a logistic regression model; however, other methods in machine learning such as boosted regression models (McCaffrey et al., 2004) have been used to estimate the propensity score

from the observed data. Apart from the PS matching and PS weighting, other approaches such as the PS stratification (Rosenbaum & Rubin, 1984) and covariate adjustment on the PS can also be used to control for confounder bias. In addition, extensions have been done on the IPW, and Rubin and Thomas (1996) proposed using the doubly robust IPW estimator that utilises a regression model on the outcome. This estimator is known to produce smallest large sample variance of any weighting-type estimator when both the propensity score model and the outcome model are correctly specified (Robins et al., 1994). Some extensions to the standard propensity score matching include full matching (Rosenbaum, 2002), covariate balance on the generalised propensity score (Imai & Ratkovic, 2014), and coarsened exact matching (CEM) (Iacus et al., 2012). It is hoped that the material presented in this chapter and the accompanying application will enable the broader use of propensity score methods in observational child and maternal health research studies in sub-Saharan Africa to mitigate confounder biases, thus allowing for assessing causal effects on exposures and interventions.

5 Future Work

We recommend future work to explore different methods of estimating the propensity score using causal inference methods in machine learning such as boosted regressions to assess the effect of interventions or risk factors on a public health outcome using observational data. Future work can consider using methods that are less sensitive to the number of type of variables is to consider for matching such as a dimension reduction technique using hierarchical clustering or principal component analysis where individuals with similar characteristics are grouped into clusters based on their principal components, and the treatment–outcome relationship can be estimated based on these clusters. The clusters identified can be considered as a proxy indicator of unmeasured confounders of the causal relationship. An explanation on a similar algorithm is done by Li et al. (2016). We further recommend the use of confounder balance methods in different study designs such as longitudinal data and survival data when the outcomes such as wasting, stunting, and underweight occur simultaneously and independence between outcomes cannot be assumed. There are several other child growth measurements that are measured from a child and can be analysed to assess child growth. These include mid-upper-arm circumference (MUAC), head circumference, and body mass index-for-age. However, this chapter focused on weight-for-age, height-for-age, and weight-for-height as they are known to be the three most common child growth measurements that provide an indication of optimal child growth (Onis, 2006). Additional work can be done to assess the impact of infant and young child feeding interventions on these growth measurements. Furthermore, researchers can explore causal effects for multiple outcome data as few research has been done in this area of which we are currently exploring.

6 Further Reading

The current state of child growth indicators and prevalence of exclusive breastfeeding for Malawi and Zambia can be found in their respective Demographic and Health Survey reports Government of Malawi and ICF (2017) and Zambia Statistics Agency and ICF (2019). A detailed explanation on infant and young child feeding can be found in WHO (2003). The updated book of Hernán and Robins (2020) provides an excellent overview of and an introduction to causal inference. The general idea of propensity score methods can be found in Rosenbaum and Rubin (1983) who provides detailed explanation on how the propensity scores achieve balance between exposure groups. Rosenbaum and Rubin (1984) give a good explanation on balancing the covariate difference between exposure groups using sub-classification, and Rosenbaum (2005) provides a good explanation on assessing the effect of unmeasured confounding on the estimated causal effects. The idea of comparing the different propensity score methods has been explained excellently by Austin and Mamdani (2006) and Austin (2011a). In this chapter, we discuss the propensity score matching and weighting for a data that is not hierarchical; however, most observational data are multilevel in nature, and hence, the standard propensity score is estimated at an individual level and ignores the clustering. Li et al. (2013), Arpino and Cannas (2016), and Arpino and Mealli (2011) provide a detailed explanation on propensity score matching and weighting for multilevel observational data. This chapter considers a selection of variables to include in the propensity score model under the strongly ignorable assumption.

Acknowledgments This work was supported through the DELTAS Africa Initiative, Sub-Saharan Africa Consortium for Advanced Biostatistics Training (SSACABT). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [SSACABT] and the UK government. The views expressed in this publication are those of the author(s) and not those of AAS, NEPAD Agency, Wellcome Trust, or the UK government. Many thanks to the South Africa Medical Research Council for providing a conducive environment and space to work on this chapter.

Appendix

STATA Code

This section presents the Stata code for analysis of the effect of appropriate complementary feeding on child nutritional outcomes. The same code can be applied for exclusive breastfeeding.

```
**********The code for generating the propensity score from
the exposure model, e.g., appropriate complementary feeding (approp)
```

```
*****************************

logistic approp i.child_sex i.mothers_education i.mothers_age
i.child_age i.residence i.wealth i.antental_visits i.hiv

predict ps

************The code in Stata for matching the covariates based on
the propensity score, e.g., for stunting (stu) ********************

psmatch2 approp, outcome(stu) pscore(ps) neighbor(1) radius
caliper(0.2)

***********Extracting the Matched Pairs***************

gen pair = _id if _treated==0
replace pair = _n1 if _treated==1
bysort pair: egen paircount = count(pair)
drop if paircount !=2
tab paircount
tab pair

*****************Estimating the effect of appropriate comp feeding on
the outcomes, e.g., stunting (stu)**

clogit stu i.approp, group(pair)

************Using the Stata package psweight to estimate the effects
using  inverse probability weighting**

psweight ipw approp stu
psweight call balanceresults()

*********************************************************
```

References

Aakvik, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics, 63*(1), 115–143.

Abadie, A., & Imbens, G.W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica, 74*(1), 235–267.

Abadie, A., Drukker, D., Herr, J.L., & Imbens, G.W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal, 4*(3), 290–311.

Ali, Z., Saaka, M., Adams, A.G., Kamwininaang, S.K., & Abizari, A.R. (2017). The effect of maternal and child factors on stunting, wasting, and underweight among preschool children in northern Ghana. *BMC Nutrition, 3*(1), 31.

Arpino, B., & Cannas, M. (2016). Propensity score matching with clustered data. An application to the estimation of the impact of caesarean section on the Apgar score. *Statistics in Medicine, 35*(12), 2074–2091.

Arpino, B., & Mealli, F. (2011). The specification of the propensity score in multilevel observational studies. *Computational Statistics & Data Analysis, 55*(4), 1770–1780.

Austin, P.C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083–3107.

Austin, P.C. (2011a). An introduction to propensity score methods for reducing the effect of confounding in observational. *Multivariate Behavioural Research*.

Austin, P.C. (2011b). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Austin, P.C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine, 33*(6), 1057–1069.

Austin, P.C., & Mamdani, M.M. (2006). A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine, 25*(12), 2084–2106.

Ayisi, R., & Wakoli, A. (2014). Exclusive breastfeeding practice: its implication on nutrition status, growth and morbidity pattern among infants aged 0-6 months. *Month, 33*, 9–9.

Becker, S.O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *The Stata Journal, 7*(1), 71–83.

Becker, S.O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal, 2*(4), 358–377.

Chekol, D.A., Biks, G.A., Gelaw, Y.A., & Melsew, Y.A. (2017). Exclusive breastfeeding and mothers' employment status in Gondar town, northwest Ethiopia: A comparative cross-sectional study. *International Breastfeeding Journal, 12*(1), 27.

Cole, S.R., & Hernán, M.A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology, 168*(6), 656–664.

d'Agostino, R.B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*(19), 2265–2281.

Dehejia, R.H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.

Disha, A., Rawat, R., Subandoro, A., & Menon, P. (2012). Infant and young child feeding (IYCF) practices in Ethiopia and Zambia and their association with child nutrition: analysis of demographic and health survey data. *African Journal of Food, Agriculture, Nutrition and Development, 12*(2), 5895–5914.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, 1231–1236.

Government of Malawi and ICF (2017) Malawi demographic and health survey 2015-16 report. Tech. rep., National Statistics Office Malawi and ICF.

Guo, S., Barth, R.P., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review, 28*(4), 357–383.

Hernán, M.A., & Robins, J.M. (2020). Causal inference. what if.

Holland, D. (1986). *Causal inference using potential outcomes: design, modeling, decision*. American Statistical Association.

Horvitz, D.G., & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association, 47*(260), 663–685.

Iacus, S.M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1–24.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 243–263.

Kamenju, P., Liu, E., Hertzmark, E., Spiegelman, D., Kisenge, R., Kupka, R., Aboud, S., Manji, K.P., Duggan, C., & Fawzi, W.W. (2017). Nutritional status and complementary feeding among HIV-exposed infants: a prospective cohort study. *Maternal & Child Nutrition, 13*(3), e12358.

Kassa, T., Meshesha, B., Haji, Y., & Ebrahim, J. (2016). Appropriate complementary feeding practices and associated factors among mothers of children age 6–23 months in southern Ethiopia, 2015. *BMC Pediatrics, 16*(1), 131.

Keele, L. (2010). *An overview of Rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data* (pp. 1–15). OH: White Paper Columbus.

Kuchenbecker, J., Jordan, I., Reinbott, A., Herrmann, J., Jeremias, T., Kennedy, G., Muehlhoff, E., Mtimuni, B., & Krawinkel, M. (2015). Exclusive breastfeeding and its effect on growth

of Malawian infants: results from a cross-sectional study. *Paediatrics and International Child Health, 35*(1), 14–23.

Kumar, A., & Singh, V.K. (2015). A study of exclusive breastfeeding and its impact on nutritional status of child in EAG states. *An International Journal of Statistics Applications & Probability, 4*(3), 435–445. https://doi.org/10.12785/jsap/040311

Kurth, T., Walker, A.M., Glynn, R.J., Chan, K.A., Gaziano, J.M., Berger, K., & Robins, J.M. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *American Journal of Epidemiology, 163*(3), 262–270.

Lee, B.K., Lessler, J., & Stuart, E.A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine, 29*(3), 337–346.

Lenis, D., Ackerman, B., & Stuart, E.A. (2018). Measuring model misspecification: application to propensity score methods with complex survey data. *Computational Statistics & Data Analysis, 128*, 48–57.

Li, F., Zaslavsky, A.M., & Landrum, M.B. (2013). Propensity score weighting with multilevel data. *Statistics in Medicine, 32*(19), 3373–3387.

Li, S., Vlassis, N., Kawale, J., & Fu, Y. (2016). Matching via dimensionality reduction for estimation of treatment effects in digital marketing campaigns. In *IJCAI* (pp. 3768–3774).

Lunceford, J.K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine, 23*(19), 2937–2960.

Manda, S.O., Kandala, N.B., & Ghilagaber, G. (2014). Advanced techniques for modelling maternal and child health in Africa. In *Advanced techniques for modelling maternal and child health in Africa* (pp. 1–7). Springer.

McCaffrey, D.F., Ridgeway, G., & Morral, A.R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9*(4), 403–425.

Normand, S.L.T., Landrum, M.B., Guadagnoli, E., Ayanian, J.Z., Ryan, T.J., Cleary, P.D., & McNeil, B.J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of Clinical Epidemiology, 54*(4), 387–398.

Onis, M. (2006). WHO child growth standards based on length/height, weight and age. *Acta Paediatrica, 95*(S450), 76–85.

Pan, W., & Bai, H. (2015). Propensity score interval matching: using bootstrap confidence intervals for accommodating estimation errors of propensity scores. *BMC Medical Research Methodology, 15*(1), 1–9.

Perkins, J.M., Jayatissa, R., & Subramanian, S. (2018). Dietary diversity and anthropometric status and failure among infants and young children in Sri Lanka. *Nutrition, 55*, 76–83.

Robins, J.M., Rotnitzky, A., & Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association, 89*(427), 846–866.

Rosenbaum, P.R. (2002). Observational studies. In *Observational studies* (pp. 1–17). Springer.

Rosenbaum, P.R. (2005). Sensitivity analysis in observational studies. *Statistics Reference Online, 4*, 1809–1814.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P.R., & Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516–524.

Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.

Rosenbaum, P.R., & Small, D.S. (2017). An adaptive Mantel–Haenszel test for sensitivity analysis in observational studies. *Biometrics, 73*(2), 422–430.

Rubin, D.B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.
Rubin, D.B., & Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 249–264.
Setoguchi, S., Schneeweiss, S., Brookhart, M.A., Glynn, R.J., & Cook, E.F. (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and Drug Safety, 17*(6), 546–555.
Twabi, H.S., Manda, S.O., & Small, D.S. (2020). Assessing the effects of maternal HIV infection on pregnancy outcomes using cross-sectional data in Malawi. *BMC Public Health*
Walters, C.N., Rakotomanana, H., Komakech, J.J., & Stoecker, B.J. (2019). Maternal determinants of optimal breastfeeding and complementary feeding and their association with child undernutrition in Malawi (2015–2016). *BMC Public Health, 19*(1), 1503.
WHO. (2003). *Global strategy for infant and young child feeding*. World Health Organization.
WHO, et al. (2006). *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. World Health Organization
WHO, et al. (2009a). *The importance of infant and young child feeding and recommended practices*. Switzerland: Geneva
WHO, et al. (2009b). *Infant and young child feeding: model chapter for textbooks for medical students and allied health professionals*. Geneva: World Health Organization.
WHO, et al. (2010). *Indicators for assessing infant and young child feeding practices: part 2: measurement*
WHO, et al. (2017). Global targets 2025 to improve maternal, infant and young children nutrition. Tech. rep., World Health Organization
Woldeamanuel, B.T., & Tesfaye, T.T. (2019). Risk factors associated with under-five stunting, wasting, and underweight based on Ethiopian demographic health survey datasets in Tigray region, Ethiopia. *Journal of Nutrition and Metabolism* 2019.
Zambia Statistics Agency MoHZ and ICF. (2019). Zambia demographic and health survey 2018. Tech. rep., Zambia Statistics Agency.
Zhang, Z. (2017). Propensity score method: a non-parametric technique to reduce model dependence. *Annals of Translational Medicine, 5*(1).

Part II
Systematic Review and Statistical Meta-Analysis

Evidence-Informed Public Health, Systematic Reviews and Meta-Analysis

Samuel A. Abariga, Michael McCaul, Alfred Musekiwa, Eleanor Ochodo, and Anke Rohwer

Abstract Evidence-based public health (EBPH) ensures that decisions about the health of a population are informed by the best available research evidence, taking into account the expertise of public health practitioners as well other factors linked to the characteristics and the context of the population. Systematic reviews are essential for EBPH decision-making, as they are designed to present the available evidence in a holistic, transparent and systematic way. In this chapter, we explain what systematic reviews are; the process of meta-analysis and its use in systematic reviews of interventions; the use of meta-analyses in systematic reviews

S. A. Abariga
Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY, USA

M. McCaul (✉)
Division Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

South African GRADE Network, Cape Town, South Africa
e-mail: mmccaul@sun.ac.za

A. Musekiwa
School of Health Systems and Public Health, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

E. Ochodo
Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

A. Rohwer
Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_5

of diagnostic test-accuracy studies; network meta-analysis and how to synthesise results when meta-analysis is not possible. We also provide a short overview of assessing the certainty of evidence using the GRADE approach; and a table of useful resources.

Keywords Evidence-based Public Health · Research Synthesis · Systematic Review · Meta-analysis · Network Meta-analysis · Diagnostic Test Accuracy · GRADE

1 Introduction to Evidence-Based Public Health, Systematic Reviews and Meta-Analysis

Anke Rohwer

1.1 What Is Evidence-Based Public Health?

Public health practitioners need to make decisions on the health and well-being of a population on a daily basis. They are faced with numerous questions such as the burden of disease in a community, the risks for developing a disease or the effectiveness of interventions to prevent a disease. While the answers to some of these questions might be evident, others might not and require critical thinking and careful consideration of existing research, input from various stakeholders and experts on the topic, characteristics and values of a population as well as various other economic and social factors. This process of decision-making is referred to as evidence-based public health (EBPH). EBPH has been defined as the 'conscientious, explicit and judicious use of current best evidence in making decisions about the care of communities and populations in the domain of health protection, disease prevention, health maintenance and improvement' (Jenicek, 1997), and the process has similarly been described as 'integrating the best available evidence with the knowledge and considered judgements from stakeholders and experts to benefit the needs of a population' by the European Centre for Disease Prevention and Control (ECDPC, 2011). EBPH mirrors the principles of evidence-based healthcare (Dawes et al., 2005) and involves (1) phrasing clear questions related to a public health problem; (2) searching for best evidence to answer this question; (3) critically appraising the evidence for validity and interpreting the results; (4) considering applicability of the evidence and implementing the evidence in public health policy and practice; and (5) evaluating the process of EBPH as well as the newly implemented policies and programmes. This five-step process facilitates a systematic approach to the decision-making process.

The first three steps of the EBPH process aim to locate, assess and interpret the results of the best available research evidence, one of the key components of EBPH.

Table 1 Examples of different types of questions and related study designs

Type of question	Study design best suited to answer the question
What is the effectiveness of a public health intervention to treat/prevent a disease?	(Cluster) randomised controlled trial
What is the risk of acquiring a disease?	Cohort study or case-control study
What are the harms of a public health intervention?	Case-control study
What is the prevalence of a disease?	Cross-sectional study
What is the incidence of a disease?	Cohort study
How accurate is a test to identify the disease?	Diagnostic test accuracy study (cross-sectional or cohort-type accuracy study)
What are the experiences of people with the disease?	Qualitative study

It is essential that we start this process with phrasing a clear question based on a knowledge gap (step 1), as we cannot find an answer if we do not know what it is that we are asking. There are various tools that can help us to phrase a clear question, one of which, the PI(E)CO framework, considers the population (P); the intervention (I), the exposure (E) or the issue (I); the comparator (C); and the outcomes of interest (O). Furthermore, we need to consider what type of question we are asking, as this will inform the type of study that is best suited to answer our question (Table 1), keeping in mind that a systematic review of all available studies on a specific question will always be better than a single study on the question (see Sect. 1.3). For example, a systematic review of randomised controlled trials (RCTs) will answer a question on the effectiveness of an intervention, whereas a systematic review of cross-sectional studies will be able to answer a question on the prevalence of a disease. Once we have phrased a clear question, we will be able to select keywords related to our PI(E)CO elements to develop a search strategy that we can use to find evidence to answer our question (step 2). When performing the search, it is useful to select a database that indexes suitable studies. A few databases to consult as a starting point include the Cochrane Library that contains systematic reviews of interventions (including those relevant for public health), diagnostic test accuracy studies and qualitative research; Epistemonikos, a regularly updated database of all health-related systematic reviews; and Health Evidence™ containing quality rated systematic reviews related to the effectiveness of public health interventions. Once we have found a suitable study, we need to critically appraise the internal validity of the study and interpret the results (step 3). This is an essential step in the process, as we always need to consider the results of a study in light of its risk of bias (or systematic error) and cannot just trust the conclusions of a study at face value. There are various tools that can assist us to assess risk of bias, depending on the type of study we are reading. The AMSTAR 2 (Shea et al., 2017) and the ROBIS (Whiting et al., 2016) tools can be used to assess risk of bias of systematic reviews. Best

available research therefore relates to the most relevant, trustworthy and up-to-date research study that is currently available.

When applying the results of the evidence to our problem (step 4), we first need to consider whether the research is generalisable to our context. This means that we need to evaluate whether the population, the setting and the interventions included in the study are similar to our own context and whether all the important outcomes, relevant to our question, have been addressed. If we think that the evidence is applicable to our setting, we then need to consider a wide range of other factors such as the burden of disease; availability of resources; socio-economic, cultural and environmental conditions; and the local context of the community. Involving experts and other relevant stakeholders at international, national and local level, healthcare workers, the general public and policymakers is key. The best available research evidence therefore is an essential part of the decision-making process, but on its own, it is not sufficient to make a decision. Indeed, some people prefer the term 'evidence-informed decision-making' which highlights this notion.

Lastly, as public health practitioners, we need to critically reflect on the decision-making process and evaluate the implementation of public health interventions (step 5). This might lead to new questions, which will lead us back to step 1 of the EBPH process.

1.2 Why Do We Need Evidence-Informed Decision-Making?

In a perfect world, all decisions about healthcare and public health should be informed by best available research evidence, as explained above. In reality, however, decisions are often made haphazardly, based solely on experts' opinions or anecdotal evidence, and aim to fulfil short-term demands of stakeholders such as politicians or funders, without considering potential harms (Brownson et al., 1999).

Using up-to-date and trustworthy evidence, in the form of systematic reviews, to inform public health decisions is vital. Sub-Saharan Africa faces numerous health challenges related to a huge burden of infectious diseases, non-communicable diseases, road traffic injuries as well as maternal and neonatal health conditions. These challenges are exacerbated by, inter alia, a lack of resources in healthcare facilities, including staff shortages, poor access to healthcare and poor socio-economic conditions. In this context, adopting policies and practices that are beneficial and effective, as well as abandoning policies and practices that are ineffective or even harmful, is key to ensure that scarce resources are not wasted (Chinnock et al., 2005).

1.3 Overview of Research Synthesis

Research synthesis refers to the process of identifying two or more studies on the same or similar question, evaluating these studies and summarising their findings.

The aim of research synthesis can be compared to building a jigsaw puzzle: a single piece of the puzzle will only give us part of the information needed, whereas using all available pieces to build the puzzle will result in us being able to see the complete picture. The rationale behind research synthesis is built upon the notion that science is cumulative. Indeed, we can only confidently interpret results of a study in the context of results of other studies addressing the same question (Sir Iain Chalmers).

Research synthesis is important because it helps make sense of research, as different studies on the same question can yield different findings. Reviewing all existing studies in a holistic way ensures that 'cherry picking' of favourable results is avoided and that information about alternative, potentially effective interventions can be considered, leading to better accountability of public health practitioners. Furthermore, research synthesis is important as it helps us to cope with information overload since we can read a single source to stay up to date with new developments instead of several individual studies. Lastly, research synthesis is a valuable tool to identify gaps in existing evidence and to justify future research, which helps to reduce research waste and reduces costs for unnecessary research.

1.4 Types of Research Synthesis

Research synthesis is an umbrella term for any kind of review or summary of the literature. There are a number of different types of reviews that differ in terms of aims and methodology (Sutton et al., 2019). A literature review is generally a qualitative, narrative summary of evidence relating to a specific topic, written by experts in the field. These 'traditional' literature reviews typically do not follow a formal process to collect and interpret information but rather rely on subjective methods. Literature reviews can be useful when introducing a new field, providing background information to a question or reviewing methods to analyse data; however, they cannot be relied on when making healthcare decisions.

Reviews that are most relevant to public health decision-making include, but are not limited to, scoping reviews, systematic reviews and rapid reviews (see Table 2).

A scoping review follows a systematic, pre-specified approach and can be undertaken to examine the extent, range and nature of the existing evidence related to a question; to inform the scope and determine the value of conducting a systematic review; to identify gaps in the evidence; or to summarise findings from various sources of evidence (Tricco et al., 2016; Tricco et al., 2018). Scoping reviews are generally descriptive in nature and are useful when exploring new areas of research or clarifying concepts.

A systematic review aims to summarise all studies relevant to a particular question in a transparent, comprehensive and rigorous way. They aim to reduce bias in the review process by having pre-specified and reproducible methods, assessing risk of bias for all included studies, synthesising results in a meta-analysis where possible and making conclusions based on the totality and the quality of evidence.

Table 2 Examples of reviews relevant to public health in sub-Saharan Africa

Type of review	Objectives
Scoping review (Mudie et al., 2019)	'to identify the level of research output on NCDs (cardiovascular disease, diabetes, obesity, respiratory disease, cancer, and chronic kidney disease), as well as their risk and prognostic factors, from large NCDs cohort studies in SSA' 'to identify any limitations and gaps and inform future research'
Systematic review (De Buck et al., 2017)	To examine the 'effectiveness of different approaches for promoting handwashing and sanitation behaviour change, and factors affecting implementation, in low and middle-income countries'
Rapid review (Zhen et al., 2020)	'to assess the abilities of different interventions to decrease the incidence of droplet-based infections among people using public ground transport'

Systematic reviews play a significant role in informing healthcare decisions (Moher et al., 1999; Moher et al., 2009).

A rapid review is a review where some of the steps of a systematic review have been omitted to make the information available in a short period of time. Rapid reviews are generally conducted in response to requests from decision-makers. There is no one-size-fits-all approach for rapid reviews, and authors need to tailor their methods according to the scope of the question and the urgency of the request (Tricco et al., 2015). During the COVID-19 pandemic, rapid reviews played a vital role in informing decisions about healthcare as there were a myriad of questions that needed to be answered and new evidence emerged daily.

1.5 What Is So Special About Systematic Reviews?

When making decisions about healthcare and public health, systematic reviews are considered best available evidence. Key features of systematic reviews include explicit, pre-specified and systematic methods (see Table 3) that are reproducible and transparent; clearly phrased, focussed questions; comprehensive search strategies that cover a variety of sources and include sources of ongoing and unpublished studies; explicit and pre-specified selection criteria that are applied equally to all potentially included studies; standardised extraction of relevant data across included studies; assessment of risk of bias according to the type of study included in the review, applied to each included study; statistical pooling of data in a meta-analysis where possible; grading of the certainty or quality of the evidence per outcome; and formulating of conclusions based on the totality of evidence and the confidence we have in the results.

Table 3 Steps of a systematic review

1. Identify a gap in the evidence-base and provide the rationale for the review
2. Formulate the systematic review question
3. Define eligibility criteria for including studies using the PICO framework
4. Develop a comprehensive search strategy
5. Develop the rest of the protocol including methods on selecting studies, assessing risk of bias and analysing data
6. Register and/or publish the protocol of the systematic review
7. Search electronic databases
8. Screen titles and abstracts of search results
9. Obtain full texts of potentially eligible studies
10. Screen full texts and determine included and excluded studies
11. Provide reasons for excluding studies
12. Extract data from included studies
13. Assess risk of bias of included studies
14. Synthesise data either by pooling data in a meta-analysis or narrative synthesis
15. Assess the certainty of evidence per outcome using GRADE
16. Interpret results
17. Write-up results and publish review
18. Disseminate findings

Systematic reviews can make use of meta-analysis, the process of statistically synthesising (sometimes referred to as 'pooling') the effects of individual studies that are deemed similar enough to be combined, to produce an overall effect estimate (see Sect. 2). A meta-analysis is usually presented graphically in the form of a forest plot. Meta-analysis should always be applied within the context of a systematic review that used rigorous methods to ensure that all eligible studies were included, that studies were homogenous enough to be combined and that results from the meta-analysis were interpreted in light of the risk of bias of included studies. Applying this technique outside of this context can be potentially dangerous and misleading.

1.6 The Value of Systematic Reviews in Public Health Decision-Making

Systematic reviews are powerful tools for public health decision-making. Meta-analysis within the context of a systematic review provides a pooled effect estimate based on the results of all included studies.

To illustrate this, imagine that you are a public health officer, responsible for distribution of insecticide-treated bed nets (ITNs). You know that some people receive the nets, but either do not use them or do not use them properly. Then there is the added concern about resistance to the insecticide that you have read

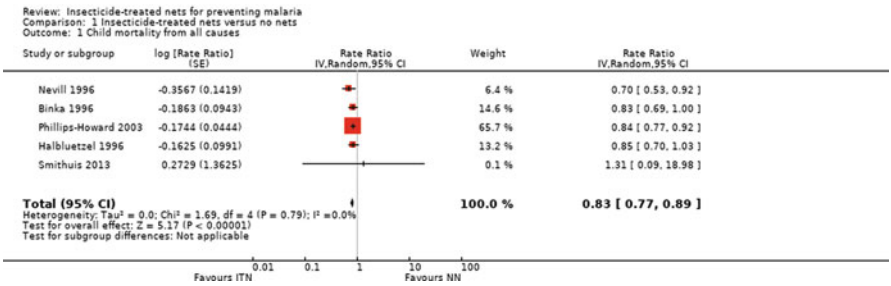


Fig. 1 Forest plot showing results for child mortality. (Reproduced from Pryce et al. (2018) under the terms of the Creative Commons Attribution Non-Commercial License CC BY-NC)

about. You wonder whether this is still an effective intervention to reduce deaths due to malaria, especially among children, who are more likely to die from malaria compared to adults. You have previously attended training on EBPH and know that you need to find research evidence to help answer your question. You do a search on PubMed and find a study by Smithuis et al. (2013), which seems to be the most recent study on this question. This cluster RCT included 8175 children, of which 26 in the ITN group and 20 in the group that did not receive bed nets died during the course of the study. This translates to a mortality rate ratio of 1.31 (95% confidence interval (CI) 0.09 to 18.98), which roughly means that we do not know whether ITNs have an effect on child mortality or not. This result first has you worried, but then you remember that a systematic review might give you a better answer to your question, since it takes into consideration all relevant studies on this question. You thus proceed to conduct a search on the Cochrane Library and find a review on 'Insecticide treated nets for preventing malaria' (Pryce et al., 2018). This review examined the effectiveness of ITNs in reducing malaria prevalence, morbidity and mortality. It included 23 RCTs with over 275,793 participants. The updated review found high-certainty evidence that ITNs, compared to no bed nets, reduce all-cause mortality in children by 17% (rate ratio 0.83, 95% CI 0.77 to 0.89). You scroll down to find the forest plot related to this finding (Fig. 1) and see that it includes the results of the cluster RCT you read initially (Smithuis et al., 2013). You realise that the result of this single RCT, which showed no definite effect, can be misleading if it is not considered in the context of other studies that have answered the same question. You make a mental note to start with a search for systematic reviews next time you need an answer to a question.

Indeed, in 2004, a Cochrane review assessing the impact of ITNs on malaria found that ITNs were effective in reducing malaria prevalence, morbidity and mortality compared to no ITNs (Lengeler, 2004). Based on these findings, there was a massive scale-up of ITN distribution, with significantly more people sleeping under a bed net in 2018 compared to 2010 (WHO, 2019b). In 2018, the Cochrane review was updated (Pryce et al., 2018) and included 23 RCTs. The review found high-certainty evidence that ITNs compared to no nets reduce childhood mortality,

uncomplicated episodes of malaria (*P falciparum*) and severe malaria episodes. This evidence informed the WHO guideline on Malaria Vector Control (WHO, 2019a) that still strongly recommends ITNs as a core intervention for populations in high-endemic areas.

1.7 Conclusion

Is it essential that decisions about the health of a population are informed by the best available research evidence, taking into consideration the expertise of public health practitioners as well other factors linked to the characteristics and the context of the population. It is promising to see that there has been increased awareness of evidence-based healthcare and public health as well as adoption of this process as part of the decision-making process in Africa in the past two decades (Young et al., 2017).

Systematic reviews are a vital part of the EBPH process. In this chapter, we explain the process of meta-analysis and its use in systematic reviews of interventions; the use of meta-analyses in systematic reviews of diagnostic test accuracy studies; network meta-analysis; and how to synthesise results when meta-analysis is not possible. We also provide a short overview of assessing the certainty of evidence within a systematic review and a table of a few useful resources.

2 Methods of Data Synthesis

2.1 Meta-Analysis

Samuel A. Abariga • Alfred Musekiwa

Meta-analysis is the statistical combination of two or more separate studies to yield an overall effect estimate along with its confidence intervals (Sacks et al., 1987). One important advantage of combining studies in a meta-analysis is to increase statistical power and improve precision of the overall effect estimates, since individual studies on their own may be too small to provide sufficient information about an intervention effect (Sacks et al., 1987; Ioannidis & Lau, 1999). When there is no substantial variability among included studies to preclude a meta-analysis, a two-step approach is used to statistically combine results. The first step is to compute a summary statistic of each included study such as the risk ratio for dichotomous data or the mean difference for continuous data. Secondly, the summary statistics of the individual studies are then combined into a single overall effect estimate, which is calculated as a weighted average of the individual studies. The weight assigned to each study depends on whether a fixed-effect or random-

effects model is applied. For the fixed-effects model, the weight is the inverse of the variance (standard error squared (SE^2)) of the effect estimate. Thus, large studies with smaller SE will receive more weight than smaller studies with larger SEs. The weighted average can be derived using the formula:

$$\text{Weighted average} = \text{sum}\left(\text{effect estimate x weight}\right)/\text{sum of weights} = \frac{\sum Yi\,Wi}{\sum Wi} \tag{1}$$

where Y_i = effect estimate of the i^{th} study and W_i is the weight assigned to the i^{th} study.

The weight assigned to each study under the inverse variance scheme is simply given as $W_i = \frac{1}{V_i}$, where V_i is the within-study error variance for individual studies.

The simplest statistical method of conducting meta-analysis of both dichotomous and continuous data is the use of the inverse variance method. Applying weights to studies decreases the variability and the imprecision of the overall effect estimates (Egger et al., 2008).

There are two main statistical methods used to perform meta-analysis: the fixed-effect and the random-effects model. The fixed-effect model assumes that all studies are estimating a common, fixed or equal effect and only considers the variability of the results within the individual studies and ignores any between-study variability (Egger et al., 2008).

A fixed-effect meta-analysis can be implemented using the generic inverse variance method by computing a weighted average of the treatment effect estimates Y_i and a summation of the individual effect estimates Y_i and weighting these by the reciprocal of their squared standard errors (SE_i) (Egger et al., 2008).

$$generic\ inverse - variance\ weight\ average = \frac{\sum Y_i\left(\frac{1}{SE_i^2}\right)}{\sum \frac{1}{(SE_1^2)}}. \tag{2}$$

A random-effects model, on the other hand, assumes that the treatment effects involving the individual studies that follow a normal distribution are assumed to vary around some overall average treatment effect (DerSimonian & Laird, 1986). Therefore the random-effects model takes into account both the within- and between-study variability in the effect estimate and weighs studies using the within- and the between-study variance (DerSimonian & Laird, 1986; Higgins et al., 2009).

Hence for a random-effects model, the observed variance for any value (Y_i) about the overall mean μ is V_i plus τ^2, i.e. ($V_i + \tau^2$), where τ^2 is the between-study variability, also known as the between-study heterogeneity. Therefore, the weight given to each study under the inverse variance will then be given as

$$W_i^* = \frac{1}{V_i + \tau^2}$$

where W_i^* is the random-effects weight. Therefore, the combined effect of the weighted mean for K included studies under the random-effects model is given by

$$\text{Weighted mean} = \frac{\sum_{i=1}^{K} W_i^* Y_i}{\sum_{i=1}^{K} W_i^*}$$

When there is no between-study variance, τ^2 will be equal to zero, and the random-effects model approximates the fixed-effect model (Borenstein et al., 2010). The assumption of no variability among included studies (homogeneity) that informs the choice of a fixed-effect model is often flawed in biomedical research given the numerous sources of diversity in study characteristics such as participants, interventions, comparators, outcomes assessed as well as study designs that are inherent with these types of research (Higgins et al., 2009). Therefore, the random-effects model is preferred in meta-analysis especially when there is high heterogeneity and can be implemented in Review Manager (Review Manager (RevMan), 2020) or any statistical package that has the capability of performing meta-analysis.

2.2 Unit of Analysis Issues

When performing a meta-analysis, it is important to keep track of the unit of randomisation of primary studies. Although this is not an issue when the unit of randomisation is the participant, it can pose a problem where the units of randomisation are groups of individuals instead of individuals themselves. Unit of analysis issues can also occur when there are more than two treatment arms. An example of a situation where unit of analysis issues can arise in meta-analysis is the inclusion of studies with cluster randomised designs. Cluster randomised studies are those in which the intervention of interest is assigned to a group (e.g. a school or clinic) as opposed to the individual and the units of observation are members of those groups (e.g. students or patients) (Murray, 1998). Clustering also occurs, for instance, where one individual contributes more than one measurement corresponding to multiple body parts such as the eyes or teeth in vision or dentistry studies, respectively. For instance, in vision research, where intervention trials are conducted such that both eyes are randomised to the same treatment group (two-eye design) or one eye is randomised to one intervention and the fellow eye to the other intervention (paired-eye design) (Murdoch et al., 1998), clustering occurs within the participant in the same way that students or patients can be clustered within a school or clinic if the school or clinic is the unit of randomisation. Other design issues that can lead to unit of analysis issues are cross-over trials, longitudinal trials in which an individual contributes more than one measurement over time. When there are more than two treatment arms, care should be taken to avoid double counting comparing more than one treatment group to the same control group. A unit of analysis error

occurs when investigators fail to account for the effect of clustering design in their analysis. Unit of analysis error also arises in meta-analysis if the included primary studies failed to account for the effect of clustering or when multiple treatment arms are not handled properly leading to double counting of participants (Whiting-O'Keefe et al., 1984).

2.2.1 Cluster Randomised Trials

Failure to account for the effect of clustering or correlation between multiple body parts from the same participant (e.g. in paired-eye design), in the analysis, may lead to overestimation of the treatment effect with a false increase in precision (Murdoch et al., 1998). The inclusion of these trials in a meta-analysis, by extension, may also lead to artificially precise estimates. One approach in dealing with this is for review authors to either elect to include only studies that accounted for clustering or correlation between multiple body parts (and provide a rationale for this decision) or correct data from primary studies that failed to account of cluster designs. Data from cluster design studies can be corrected by reducing the sample size of each trial to its approximate effective sample size (Rao & Scott, 1992). The effective sample size of each intervention group in a cluster-randomised trial is derived by dividing its original sample by the design effect of clustering. The design effect is given in the following example using the formula (Higgins et al., 2019b):

- Design effect = 1 + ([average cluster size)-1] * intra-cluster (or intraclass) correlation coefficient (ICC).
- For example, in a cluster randomised trial of which 12 hospitals with 200 patients (hospital A) were randomised into an intervention group and 13 hospitals with 250 patients (hospital B) randomised into a control group, assuming that for a dichotomous data, the trial investigators failed to account for cluster effect, if events in hospital A are 40/200 and those in hospital B are 50/250.
- The average cluster size is then computed as (200 + 250)/(12 + 13) = 450/25 = 18.
- Assuming an ICC of 0.02 (based on estimates from prior studies in the area), the design effect is given as 1 + ([18–1]) * 0.02 = 1.34.
- The effective sample size in the intervention and control group as well as the event rates in both groups must be computed, taking into account the design effect of (1.34), as follows:
- Intervention group (effective sample size) = 200/1.34 = 149.3
- Control group (effective sample size) = 250/1.34 = 186.6
- Similarly, the effective rate of events accounting for the design effect will be given by:
- Effective event in the intervention group = (40/1.34)/149.3 = 30/149.3
- Effective event in the control group = (50/1.34)/186.6 = 37/186.6
- These approximated corrected values can then be used for meta-analysis.

The 'inflated standard error' approach is the second (equivalent) approach that can be used to adjust for clustering effect. This approach is more common and

preferable as it eliminates the need to round up the effective sample size to whole numbers and can be used to adjust for clustering effect in both categorical and continuous data. The inflated variance can be used to perform the meta-analysis. To compute the inflated standard error, we first compute design effect as described above. Second, we ignore the clustering effect and compute the standard errors from confidence intervals of the effect estimate. Third, we multiply the computed standard errors, as if there was no clustering, by the square root of the design effect. This gives a new quantity called the inflated standard error, which can then be used along with the log odds ratio to perform meta-analysis using the generic inverse-variance method (Higgins et al., 2019b).

2.2.2 Studies with Multiple Arms

Multi-arm trials with a single or multiple intervention and/or comparator arms may be encountered when conducting meta-analysis. In this situation, several approaches can be explored to include the arms and avoid double counting of participants. One approach is to focus only on the intervention and comparator relevant to the review (those that meet criteria for inclusion in a pairwise comparison of intervention) and ignore the arm(s) that are not relevant to the review. A second approach is to combine any multiple interventions and/or comparator arms into a single pairwise comparison. A third approach is to divide the 'common' comparator arm equally according to the number of the intervention arms. All these approaches are aimed to prevent double counting of participants and a unit of analysis error assuming the comparisons are independent (Higgins et al., 2019b).

2.3 Assessment of Heterogeneity

Statistical heterogeneity can be assessed informally by inspecting forest plots for overlap of confidence intervals of the effect estimates. Poor overlap is an indication of the presence of statistical heterogeneity. A statistical evaluation of heterogeneity tests the null hypothesis of homogeneity, i.e. all the studies evaluate the same effect. The chi-squared (χ^2) and the Cochran's Q are used to evaluate whether differences between results are due to chance alone. Cochran's Q is calculated by the sum of the squared deviation of each individual's study estimate from the pooled meta-analysis estimate and weighing each study contribution as in the meta-analysis (Higgins et al., 2003).

$$Q = \sum \frac{1}{SE_i^2} \left(Y_i - Y_{pooled}\right) 2$$

where Y_i is the individual study effect estimate and Y_{pooled} is the pooled meta-analysis estimate. For k included studies, a p value is computed by comparing the

χ^2 statistic with *k-1* distribution of freedom (Higgins et al., 2003). Because the test has low power at identifying true heterogeneity, to offset for the low power, a cut-off of ($P < 0.1$) is used to indicate statistical significance (Dickersin & Berlin, 1992). The extent of the presence of statistical heterogeneity in the pooled meta-analysis estimates is denoted as I-square (I^2). The I^2 describes the percentage of the total variability that is explained by heterogeneity and not due to chance (Higgins & Thompson, 2002; Higgins et al., 2003). I^2 can be computed as:

$$I^2 = 100\%X\,(Q - df)\,/Q$$

where *Q* is Cochran's heterogeneity statistic and df the degrees of freedom. I^2 ranges between 0% and 100%. I^2 range from 0% to 40% indicates that the presence of heterogeneity may not be important, and a range between 30% and 60%, 50% and 90% and 75% and 100% indicates moderate, substantial and considerable statistical heterogeneity, respectively (Deeks et al., 2019). The presence of substantial or considerable statistical heterogeneity may be a reason for deciding not to perform a meta-analysis (Higgins et al., 2003).

2.4 Methods for Dealing with Heterogeneity in Meta-Analysis

When a meta-analysis results in significant statistical heterogeneity, there is a need to investigate the sources of heterogeneity. An eyeball inspection of the forest plot can reveal one or more outliers, which refers to studies having extreme effect estimates compared to most studies included in the meta-analysis. The first step should be to check for any errors in the data extraction because this could be the reason for the outliers. After establishing that there are no errors in data extraction, one can either decide not to perform meta-analysis or to investigate whether there are differences between studies with respect to participants, interventions, comparisons, outcomes or study design, which may explain the heterogeneity. This is carried out using two statistical techniques, namely, subgroup analysis and meta-regression (Deeks et al., 2008).

2.4.1 Subgroup Analysis

Subgroup analysis is undertaken by separating the overall meta-analysis according to important subsets of participant or study characteristics that may interact with or modify the effect of the intervention (e.g. children versus adults) or subsets of study setting (e.g. community versus tertiary level of care), intervention route of administration (e.g. intravenous versus oral route) or study geographical region (e.g. country or continent). When doing systematic reviews and meta-analysis, subgroup analysis may be necessary to investigate the source of heterogeneity or to answer

specific questions regarding specific subgroups or assess important participant or trial characteristics that may affect the effect estimate (Deeks et al., 2008).

Subgroup analyses are decided a priori, and even if there is no significant statistical heterogeneity detected, it is advisable to still perform the subgroup analysis as the results may differ between subgroups. A Cochrane review on antibiotics for acute otitis media in children where the outcome pain was compared between antibiotic versus placebo group is used here to illustrate subgroup analysis according to the period at which pain was measured (after 24 hours, 2–3 days, 4–7 days and 10–12 days) (Venekamp et al., 2015). We used Review Manager software and data from the published Cochrane systematic review to reconstruct the forest plot showing the subgroup analysis (Fig. 2). Although the overall meta-analysis revealed some level of heterogeneity ($Chi^2 = 31.43$, degrees of freedom [df] =21, $P = 0.07$, $I^2 = 33\%$), there was very little heterogeneity for pain at 24 hours ($Chi^2 = 4.68$, $df = 5$, $P = 0.46$, $I^2 = 0\%$) and pain at 2 to 3 days ($Chi^2 = 4.82$, $df = 6$, $P = 0.57$, $I^2 = 0\%$). This may imply that the period at which pain was measured could be the source of heterogeneity. However, there was a higher level of heterogeneity for pain at 4 to 7 days ($Chi^2 = 11.65$, $df = 7$, $P = 0.11$, $I^2 = 40\%$). In terms of the summary effect, although there was significant antibiotic effect overall (risk ratio [RR] 0.78, 95%CI: 0.71 to 0.86), there was no real effect of antibiotics on pain at 24 hours (RR 0.89, 95%CI: 0.78 to 1.01). This strengthens the need for subgroup analysis as the effect of the antibiotic may have been modified by the period at which pain was measured; this is further confirmed by the significant subgroup differences ($Chi^2 = 10.77$, $df = 3$, $P = 0.01$, $I^2 = 72.1\%$).

When subgroup analysis reveals opposite intervention effect in separate subgroups, it is known as ‘qualitative’ interaction. When subgroup results are in the same direction but differ quantitatively, this is referred as ‘quantitative’ interaction. However, the Cochrane Handbook cautions against comparing results from different subgroups and also calls for careful interpretation of findings from subgroup analysis (Deeks et al., 2008).

Subgroup analysis may be hampered when studies do not present separate results for different subgroups. Suppose an analyst wants to perform subgroup analysis on sex variable but some studies do not have sex-disaggregated results. It will not be possible to include these studies in the subgroup analysis unless separate results are requested from primary study authors, which is usually unsuccessful. The solution to this problem can be individual patient data (IPD) meta-analysis (Deeks et al., 2008); however this may also suffer from the same problem as obtaining IPD may prove to be even more difficult.

Subgroup analysis may yield incorrect results as it breaks the randomisation in the primary studies and multiple subgroup analyses may also be misleading (Deeks et al., 2008). It is therefore important to interpret findings from subgroup analyses with extreme caution. It is also important to mention that to avoid reporting bias, subgroup analyses should be decided a priori at the protocol stage. Lastly, subgroup analysis results are also limited by reduced sample size and statistical power to detect differences between studies, as there will be fewer studies in each subgroup as a result of missing data, for instance.

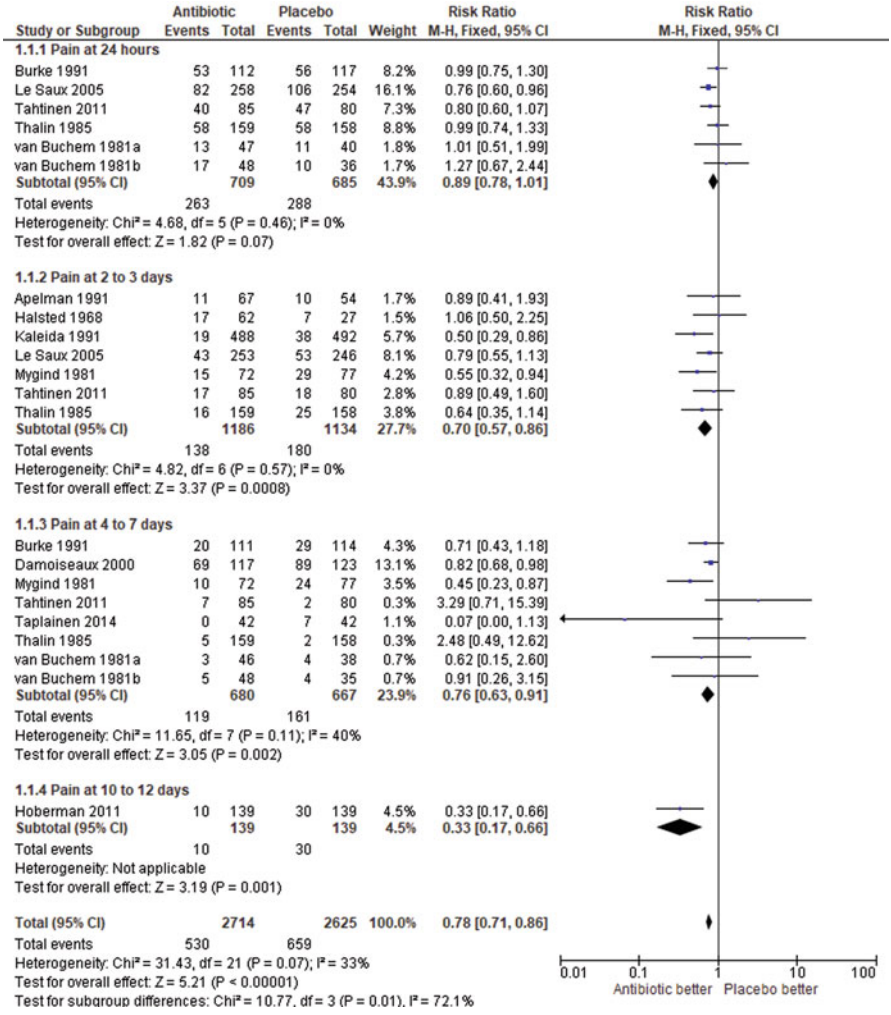


Fig. 2 An example of subgroup analysis from a Cochrane review on antibiotics for acute otitis media in children for the pain outcome. (Figure reconstructed using Review Manager with data extracted from the original Cochrane review (Venekamp et al., 2015))

2.4.2 Meta-Regression

The limitation with subgroup analysis is that it only allows one variable at a time. Meta-regression is the extension of subgroup analysis and allows investigation of multiple variables to see whether there are differences in the effect estimate between levels of variables used in the meta-regression. It is a multiple regression analysis where the outcome variable is the effect measure (can be continuous or binary) and the predictor variables are the variables that could be used for subgroup

analysis. The unit of analysis is the study, and therefore the statistical power of meta-regression is usually lower as in most cases, and the number of studies in meta-analysis is small. If there is significant association between the outcome variable and the predictor variable, then the predictor variable is a potential source of heterogeneity.

In a meta-analysis of 140 studies on prevalence of *Escherichia coli O157* (Islam et al., 2014), a meta-regression using Stata (version 13, StataCorp, College Station, TX, USA) command **metareg** was used to investigate potential sources of heterogeneity. Potential sources included world region, type of specimen, origin of sampled cattle, type of cattle, health status, pre-enrichment, immune-magnetic separation (IMIS) and isolation media. The results from the multivariable adjusted regression model found significant differences in the prevalence of *O157* with respect to world region (Africa higher compared to Asia) and type of cattle (feedlot higher compared to dairy).

However, like in subgroup analysis, the results from meta-regression need to be interpreted with caution (Deeks et al., 2008). Some studies may not provide information required to do meta-regression, for instance, there may not be separate results for different levels of the predictor variable. This effectively reduces the sample size and hence the need to interpret meta-regression results with caution.

2.5 Sensitivity Analysis

When conducting systematic reviews and meta-analysis, there are many choices that are made in selecting appropriate methods to use. These decisions may influence the results obtained in the review. Sensitivity analysis investigates whether the overall findings are sensitive or differ among all possible choices that could have been made. Variations in the statistical model used (fixed effect versus random effects), inclusion or exclusion of high risk of bias studies (for instance, in terms of allocation concealment or blinding) and different statistical analysis methods (e.g. intention to treat, missing data) can be candidates for sensitivity analysis (Deeks et al., 2008). Sensitivity analysis may also involve exclusion of some studies from meta-analysis and then comparing meta-analysis results with or without those studies.

Using data on death at final follow-up from a Cochrane review on therapeutic hypothermia for head injury, Review Manager software was used to perform sensitivity analysis to determine whether the meta-analysis results were robust to the inclusion or exclusion of studies with high risk of allocation concealment bias (Gadkary et al., 2002). The results show that although the conclusion of no significant differences between treatment and control was similar in the two meta-analyses, the effect had been overestimated by including high risk of bias studies.

Robust results where the effect estimate is the same under different assumptions or methods reinforce belief in those findings. As in subgroup analysis, the Cochrane Handbook of Systematics of Interventions cautions on careful interpretation of results from sensitivity analysis (Deeks et al., 2008).

3 Emerging Techniques in Systematic Reviews and Meta-Analysis

3.1 Network Meta-Analysis

Michael McCaul

3.1.1 Introduction to Network Meta-Analysis: The Extension to Pairwise Meta-Analysis

Most systematic reviews provide a pooled effect between two pairs of interventions (usually an intervention and control) for a specific clinical question, with defined population. However, as in public health, there might be various competing interventions available to treat a given condition or patient. Decisions often need to be made as to 'which intervention is best for my patient?' or 'Which intervention is safest or how does this intervention compare to all other options available?'. Such questions are not easily answered by traditional pair-wise meta-analysis but rather by considering an analytical approach that compares all interventions simultaneously.

Network meta-analysis (NMA) is the natural extension to traditional pairwise meta-analysis that allows the synthesis of a network of studies (such as RCTs) that compare different treatment interventions. NMA is a relatively recent development, which extends the principles of traditional meta-analysis to the evaluation of multiple treatments in a single analysis. This is possible as the connected network of studies – all answering a similar overarching question – provides both direct and most importantly indirect evidence (Bucher et al., 1997; Salanti, 2012). Traditional meta-analysis is useful, however limited, in that it can only compare two interventions at a time and only those studied directly in head-to-head trials. For example, in mental health, there are various classes of interventions that may improve health outcomes, such as psychosocial, psychological and pharmacological interventions (Mavridis et al., 2015). Even within a class of interventions, there is a range of available options that are different from each other; consider, for example, that there are up to 45 pharmacological interventions available to treat social anxiety disorder. A NMA systematic review is able to provide an answer to which intervention is the best, considering the comparative effectiveness and potential for harm of all interventions available for a particular condition.

NMA estimates treatment effects using a combination of direct and indirect evidence, called mixed evidence. Direct evidence refers to evidence that is obtained from, for example, RCTs that compare interventions B and C in a head-to-head comparison. Direct evidence is thus the relative effects between B and C. Indirect evidence is obtained through calculating the relative effects of interventions through common comparators in a network of interventions (Li et al., 2011). For example, in the absence of B and C direct (head-to-head) comparisons from RCTs, we can

estimate the relative effect of B and C via another common comparator A, forming an evidence loop (A-B-C). Thus we can obtain indirect estimates via a network of trials for B vs C from RCTs comparing A vs C and A vs B, where mathematically it can be written as, where μ is the effect estimate (e.g. mean difference) or *var* is the variance:

$$\mu_{BC}^{ind} = \mu_{AC}^{dir} - \mu_{AB}^{dir}$$

$$\text{var}\left(\mu_{BC}^{ind}\right) = \text{var}\left(\mu_{AC}^{dir}\right) + \text{var}\left(\mu_{AB}^{dir}\right)$$

When four or more competing interventions are available, the indirect estimate can be determined via multiple indirect pathways. In this case, the only requirement is that two interventions are connected (within the evidence loop) via a single common comparator (Chaimani et al., 2019).

However, just like traditional meta-analysis, NMA has various assumptions that have to be met in order for results to be valid.

3.1.2 Assumptions of NMA: Transitivity and Inconsistency

The assumption of transitivity is essential to a NMA and refers to the validity of indirect and direct assumptions. Transitivity implies – similar to the pairwise meta-analysis homogeneity assumption – that the distribution of study characteristics (effect modifiers) is similar across treatment comparisons. Conceptually, the transitivity assumption is that all interventions in the network are jointly randomisable, that is, one should be able to compare all the interventions in a single multi-arm RCT. Sources of intransitivity typically originate from clinical, population, intervention, comparison, outcome (PICO) or methodological differences between trials. For example, in a NMA of pharmacological treatments for patients with social anxiety disorder, the authors tested the assumption of transitivity through meta-regression of important effect modifiers (Williams et al., 2020). Baseline severity and mean age at onset of condition were tested across different pairwise comparisons, as any significant imbalance in these covariates across trials could signal intransitivity and undermine our confidence of the indirect evidence estimates (Wood et al., 2008). Another example includes the effect of trails conducted in different time periods as older trials typically have smaller sample sizes or are of worse quality and exaggerate effects (Chaimani et al., 2019).

The statistical manifestation of transitivity is that of the coherence (or consistency) assumption, whereby there is agreement between different sources of direct and indirect evidence. Incoherence can be tested in a closed loop of evidence (any subset of interventions where each of that have been directly compared with one another) by calculating an inconsistency factor. Several approaches have been suggested for evaluating incoherence in a NMA of interventions with various loops;

these methods are commonly known as local and global approaches (Donegan et al., 2013; Veroniki et al., 2014). Local approaches consider groups of the network separately to detect incoherence, whereas global approaches consider incoherence in the entire network. In summary, transitivity and coherence in NMA can be considered i) during protocol development, where treatments being compared are in principle jointly randomisable; ii) while looking at the studies in the NMA, where groups of studies do not differ with respect to the distribution of effect modifiers; and iii) by analysing the data, where various statistical test can determine whether direct and indirect treatment effects are in statistical agreement.

3.1.3 NMA Analysis in STATA and Presenting Results

Network meta-analysis can be performed using several approaches and via various analytical programs (Salanti, 2008; Salanti, 2012). These range from using meta-regression when there are no multi-arm trials in the connected network to using hierarchical models (via a Bayesian framework) (Sobieraj et al., 2013; Petropoulou et al., 2017) or multivariate meta-analysis methods (via a Frequentist framework) (Jackson & White, 2011; Mavridis & Salanti, 2011). STATA uses the command *mvmeta* to conduct a NMA in a Frequentist framework (White, 2011).

3.1.3.1 Network and Contributions Plots

One of the first steps in conducting a NMA is to understand the evidence base through a network graph. A network graph provides a visual representation of the network structure and is important in establishing analytical strategies and interpreting the results. Figure 3 represents an example of a network graph of pharmacological interventions from trials of adults with social anxiety disorders. It consists of nodes (circles) representing the interventions in the network, while the lines indicate direct comparisons between pairs of interventions. Depending on the network graph settings, the thickness of the lines might represent the data available for that comparison, while the size of the nodes typically indicates the number of studies included in that node. The STATA command *networkplot* can be used to create a network graph. Alternative weighting variables for nodes and edges can be made with the options *nodeweight()* and *edgeweight()*.

In this network graph (Fig. 3), placebo is clearly the reference standard, while paroxetine is the most frequent active comparator. Additionally, one can see there are four 'closed loops', indicating possible multi-arm trials. Network graphs can also be modified to indicate risk of bias in a network by using coloured lines of direct comparisons. The risk of bias in studies included in our network should affect our confidence in the quality of the direct evidence estimates and extension network estimates as stipulated by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group (Brignardello-Petersen et al., 2018). The option *edgecolor()* can change the edges of the network plot

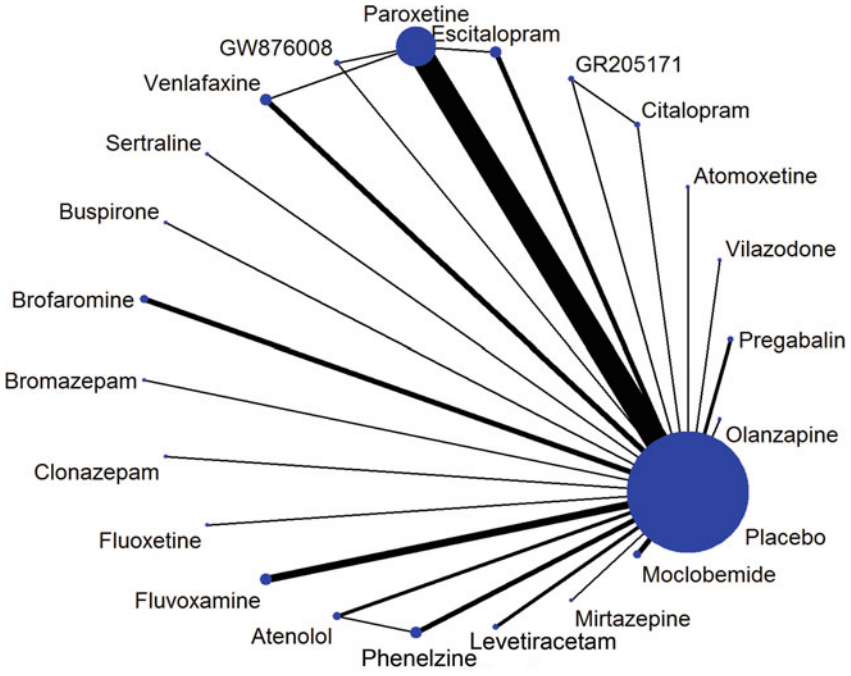


Fig. 3 Network graph of pharmacological interventions from trials of adults with social anxiety disorders. (Data to create graph taken from Williams et al. (2020))

according to the level of design limitations of included studies for each pairwise comparison. By adjusting the width of each edge to be proportional to a continuous effect modifier (e.g. year of publication or age), a visual inspection can then be useful in uncovering whether the transitivity assumption is likely to hold.

Each direct comparison in a NMA contributes differently to the network summary effect. To identify the most influential comparisons or contributions of individual treatments in a matrix plot, the command *netweight* can be used.

3.1.3.2 Performing Network Meta-Analysis

To perform a NMA in STATA use the *mvmeta* command, where first steps would be to get the data in the right format. This can be done by using the *network* package that calls the *mvmeta* command. The network package fits a network meta-analysis model, either a consistency model or a design-by-treatment interaction model (Higgins et al., 2012). The *network setup* command creates the observed treatment effect estimates and variance-covariance matrix in each trial. This *network* package includes useful syntax for getting started and creating descriptive tables and figures, various utilities and lastly analysis and graphical options. Use *network*

Table 4 League table of network estimates

	ATN	ATM	BROF	BZO
ATN	.	1.74 (0.20-14.84)	0.13 (0.04-0.47)	0.05 (0.01-0.33)
ATM	0.60 (0.07-4.92)	.	0.08 (0.01-0.61)	0.03 (0.00-0.36)
BROF	7.60 (2.12-27.22)	13.20 (1.65-105.47)	.	0.41 (0.07-2.31)
BZO	18.77 (3.04-115.67)	32.59 (2.81-377.44)	2.47 (0.43-14.01)	.

Data taken from Williams et al. (2020)

meta to estimate the relative effectiveness between each treatment and the reference standard along with standard errors, p-value and 95% confidence intervals.

The *netleague* command creates a 'league table' of all estimated pairwise summary effects. This provides a matrix of the effectiveness and uncertainty of all pairs of interventions or comparisons. For example, Table 4 provides a summary of network treatment estimates (with 95% confidence intervals) of pharmacological interventions for social anxiety disorders (extract from original table) with green highlights indicating significant effect estimates, where the upper and lower diagonals are mirrored effect estimates (odds ratios) with confidence intervals.

Using the *network forest* command creates forest plot of NMA data. Forest plots are useful to inspect the treatment effects qualitatively in considering heterogeneity and network assumptions. Forest plots should report appropriate measures of heterogeneity such as the I^2 and chi^2 for significant heterogeneity. In STATA *forest* provides network estimates and if appropriate superimposed inconsistency estimates via the inconsistency plot or model. Creating an inconsistency plot is part of evaluating and presenting assumptions of NMA.

3.1.4 Presenting and Evaluating and Assumptions of NMA

There are various attractive options available to evaluate and present the key assumptions in NMA. An inconsistency plot provides an overview of differences between the direct and indirect effect estimates for paired comparisons in the network. Violation of consistency is an important threat for the validity of results. Inconsistency should be checked in each closed loop of evidence. This is where the network graph becomes useful in identifying closed loops such as triangular (loops formed by three treatments all compared to each other) loops. This approach, termed the loop-specific approach, estimates the inconsistency factor (*IF*) as the absolute difference between direct and indirect evidence for each comparison in the loop with 95% confidence intervals. The inconsistency plot via the command *ifplot* identifies all triangular and quadratic loops in a network and provides the *IF* (see Fig. 4). Inconsistency should be evaluated by looking at each closed loop in the network separately, typically by looking at the ratio of odds ratios

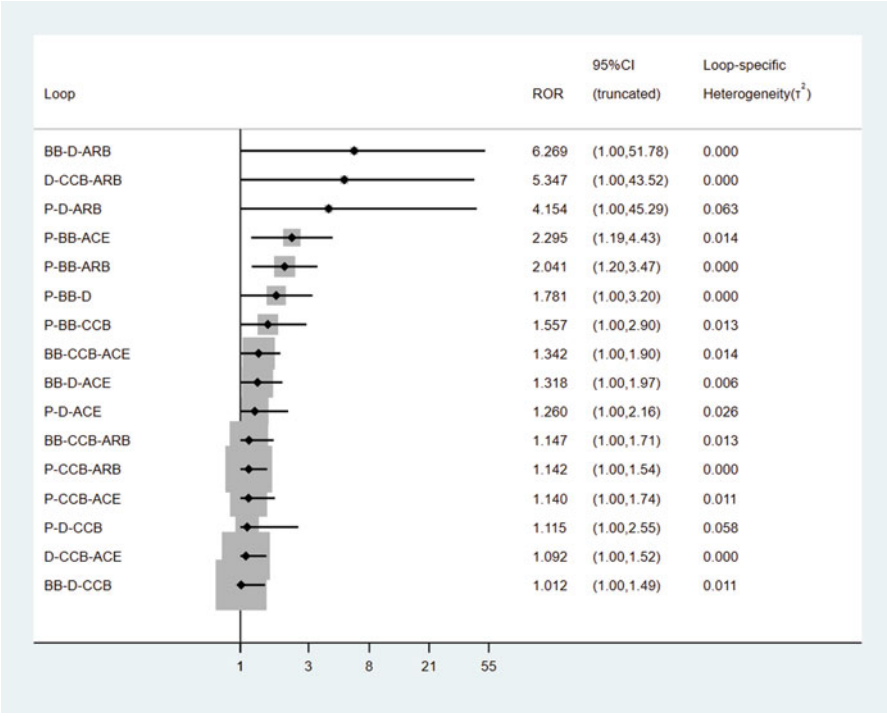


Fig. 4 Inconsistency plot in STATA. (Data to create plot taken from Elliott and Meyer (2007))

between direct and indirect estimates in every loop by plotting the *eform* of the *IFs*. Consistency is present if the *RoR* 95% confidence intervals stretch over 1 (RoR = 1). Interpret with caution if there are wide confidence intervals or large inconsistency values. Inconsistency should be explored using meta-regression (using *metareg*), or consider using statistical models that relax the assumptions of inconsistency (Cooper et al., 2009; Higgins et al., 2012; White et al., 2012). Additionally the *mvmeta* command allows for alternative methods of inconsistency testing, including the loop-specific approach described above and the design-by-treatment interaction model (differences between trials with different designs such as two-arm vs. multi-arm trials).

Another important assumption in both pairwise and network meta-analysis is that of heterogeneity. In pairwise meta-analysis visual inspection of the forest plot, together with interpretation of the I^2 measure and other statistics, helps infer about the magnitude of heterogeneity in the data. In NMA, we use a predictive interval plot to assess between-study heterogeneity using the command *intervalplot* after running *mvmeta*. The predictive interval is the interval within which the relative effects of a future study are expected to lie (i.e. the impact of heterogeneity on each comparison; see Fig. 5).

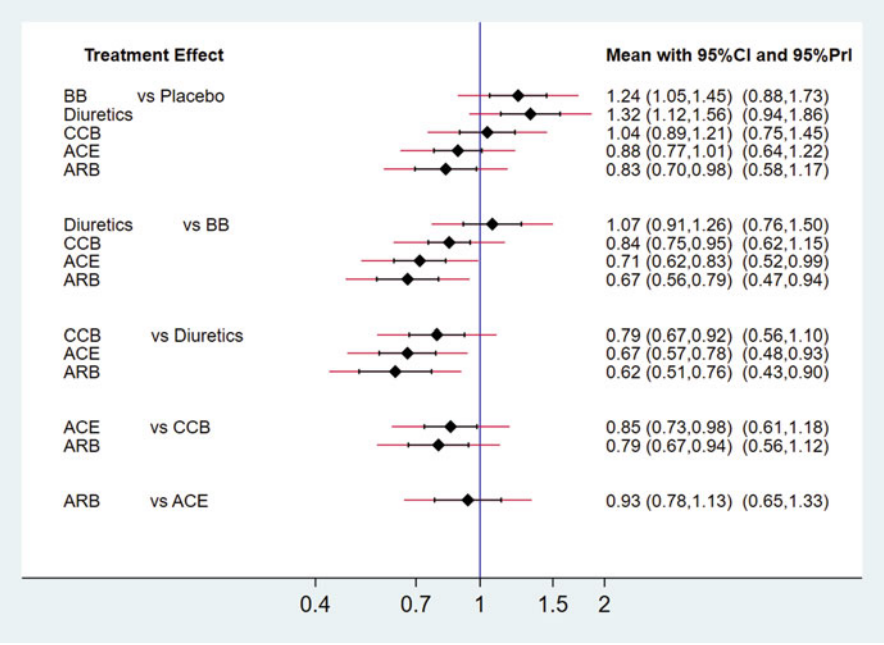


Fig. 5 Forest plot with predictive intervals. (Data to create plot taken from Elliott and Meyer (2007))

3.1.4.1 Ranking Interventions in NMA

One of the key advantages of NMA over traditional pairwise meta-analysis is the ability of NMA to rank treatment effects relative to each through estimating ranking probabilities. Ranking tables list the probability of being rank 1, and rankograms plot these values in graphs. Surface under the cumulative ranking curve (SUCRA) plots depict the cumulative ranking curve and show the distribution of ranking probabilities for each intervention. These should be interpreted with caution and always using (where possible) 95% confidence intervals. The STATA command *network rank* will provide a ranking table of interventions and *sucra* with *rankograms* as an option to produce rankograms instead of cumulative ranking curves.

3.1.5 Conclusions

Network meta-analysis allows researchers to compare all treatments available for a given condition, even if some direct comparisons are lacking, through the combination of direct and indirect information. NMA is very attractive in the public

health setting where multiple competing interventions are in play. NMA answers clinicians' and patients' questions such as which treatments are the best.

3.2 Meta-Analysis of Diagnostic Test Accuracy Studies

Eleanor Ochodo

3.2.1 Introduction to Meta-Analysis of Diagnostic Test Accuracy Studies

Accurate and effective diagnosis of disease conditions is essential to guide appropriate and timely management and treatment. The accuracy of a test is its ability to distinguish those with a disease condition or state from those without (Šimundić, 2009; van Stralen et al., 2009). Like other areas in public health, meta-analyses of diagnostic test accuracy are used to make evidence-based decisions about the implementation and interpretation of diagnostic tests (Leeflang et al., 2008; Ochodo & Leeflang, 2012; Takwoingi et al., 2015). Evidence points to their increasing use to inform guidelines and policy for clinical and public health testing (WHO, 2015, 2020a, b).

Meta-analyses of diagnostic test accuracy are useful in summarizing and obtaining precise estimates of accuracy when many small primary studies are available. They are also useful in establishing why test accuracy varies and in comparing the performance of different tests (Leeflang et al., 2008; Takwoingi et al., 2015). Test accuracy can vary due to various factors including the patient population, prevalence of disease, spectrum or disease stage, the type and quality of test, how a test is administered or conducted as well as how a test result is interpreted based on test thresholds (what defines a positive or negative test) (Whiting et al., 2004; Whiting et al., 2013).

Meta-analyses of diagnostic test accuracy studies are more complex and differ from that of intervention studies as they deal with analysis of two paired outcomes of interest, most common being sensitivity (proportion of those with a disease condition who test positive) and specificity (proportion of those without a disease condition who test negative). Sensitivity and specificity are the most common measures of test accuracy reported in primary studies, hence used in meta-analyses (Takwoingi et al., 2015).

In this section we discuss the meta-analysis of (DTA). We will introduce basic concepts of diagnostic test accuracy studies, highlight factors influencing the conduct of a meta-analysis of DTA studies and discuss methods and steps for meta-analysis of diagnostic test accuracy.

Table 5 Two-by-two table

Index test result		Reference test result	
		Positive (those with disease)	Negative (those without disease)
	Positive	TP	FP
	Negative	FN	TN

3.2.2 Basic Concepts of Diagnostic Test Accuracy

To evaluate the accuracy of a diagnostic test, the results of the index test (test under evaluation) are compared to that of a reference test (best available test) (Šimundić, 2009, van Stralen et al., 2009). These results are then cross classified in a two-by-two table to show the number of true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) (Table 5). TP, FP, FN and TN results are useful in knowing downstream consequences of a test. For example, FP results may lead to overtreatment, and FN results may lead to missing of appropriate treatment. Accuracy measures can be derived from two-by-two tables. There are two groups of accuracy measures: i) paired measures of accuracy including sensitivity and specificity, positive and negative predictive values and positive and negative likelihood ratios and ii) global measures of test accuracy, for example, diagnostic odds ratios and the area under the curve (Table 6). Global measures of accuracy are used for general assessment to evaluate the overall discriminative power of tests and to compare the overall performance of two or more tests. Paired measures relate to the test's ability to detect or exclude disease in specific populations or demonstrate the clinical significance of a positive or negative test result to a specific patient (Šimundić, 2009; van Stralen et al., 2009; Takwoingi et al., 2015).

3.2.3 When to Conduct a Meta-Analyses of Diagnostic Test Accuracy Studies

A meta-analysis of diagnostic test accuracy can be done to summarise the accuracy estimates of a single test, to compare the accuracy of two or more tests and to investigate sources of heterogeneity or variability in accuracy estimates between and within studies (Macaskill et al., 2010; Takwoingi et al., 2015).

However, just like intervention studies, a systematic review of diagnostic accuracy studies may or may not contain a meta-analysis. In general, factors influencing the conduct of a meta-analysis include the extent of variability of study results (the greater the variability, the lesser the indication for a meta-analysis) and the number of studies (e.g. some statistical packages require four or more studies for meta-analysis models to converge) (Macaskill et al., 2010; Takwoingi et al., 2015; Higgins et al., 2019a).

Table 6 Measures of test accuracy

Measure	Formula	Definition
Sensitivity	tp/(tp + fn)	The probability of a test to detect the **presence** of disease in **someone with the disease** (positivity in disease)
Specificity	tn/(fp + tn)	The probability of a test to detect the **absence** of disease in **someone without the disease** (negativity in the absence of disease)
Positive predictive value	tp/(tp + fp)	The probability that a person with a **positive** test **has the disease**
Negative predictive value	tn/(fn + tn)	The probability that a person with a **negative** test **does not have the disease**
Positive likelihood ratio (LR+)	sens/1-spec	The likelihood of a positive test result in those with disease compared to those without disease (rules in disease)
Negative likelihood ratio (LR-)	1-sens/spec	The likelihood of a negative test result in those with disease compared to those without disease (rules out disease)
Diagnostic odds ratio	(tp*tn)/(fn*fp) or LR+/LR-	The odds of positivity in those with disease compared to those without disease
Receiver operating characteristic curve (ROC curve) and area under the curve	Graphical plot of sens (y-axis) against 1-spec (x-axis)	ROC curve plots pairs of sensitivity and 1-specificity for every individual test threshold or cut-off. The area under the ROC curve estimates the discriminative ability of the test

Unlike meta-analyses of intervention studies, meta-analyses of diagnostic test accuracy do not routinely use statistical measures of variability or heterogeneity to quantify the extent of variability (Macaskill et al., 2010). When there is a univariate outcome or single measure of effect, for example, risk ratio or odds ratio in meta-analysis of intervention studies, two statistical measures are used to test and quantify the extent of variability (the Cochran Q test and I^2) (Naaktgeboren et al., 2016; Higgins et al., 2019a). The Cochran Q test is used to test for the presence of variability beyond chance, whereas the I^2 test is used to quantify this variability. Values for the I^2 test range from 0 to 100% with more than 50% representing

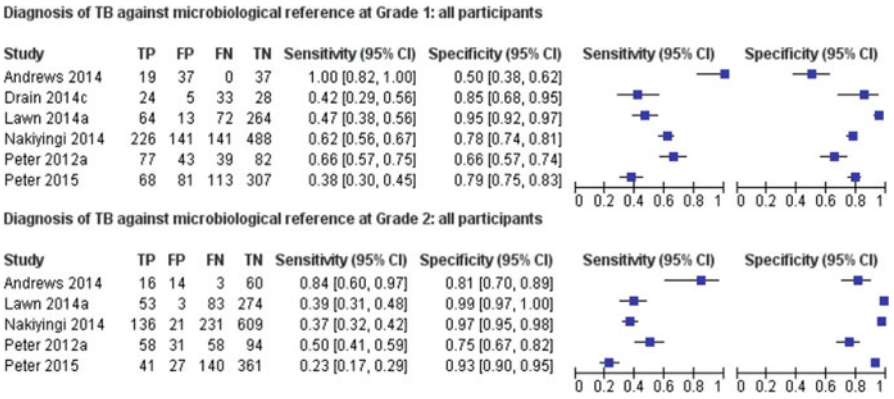


Fig. 6 Example of a forest plot for a meta-analysis of diagnostic test accuracy. (Reproduced from Shah et al. (2016) under the terms of the Creative Commons Attribution Non-Commercial License CC BY-NC)

substantial heterogeneity (Naaktgeboren et al., 2016, Higgins et al., 2019a). Paired measures of accuracy, such as sensitivity and specificity, are negatively correlated as they vary inversely with the threshold or test cut-off at which patients are labelled as diseased or non-diseased (Takwoingi et al., 2015; Naaktgeboren et al., 2016). The statistical measures (Cochran Q and I^2) do not account for variability due to test positivity thresholds and are currently not recommended for bivariate diagnostic outcomes (Takwoingi et al., 2015, Naaktgeboren et al., 2016). Instead, the Cochrane handbook for diagnostic test accuracy reviews recommends visual assessment of the variation between studies by graphically presenting the estimates of sensitivity and specificity in forest plots and summary ROC curves (Macaskill et al., 2010). An example of a forest plot for a meta-analysis of diagnostic test accuracy can be found in Fig. 6.

3.2.4 Methods for Meta-Analyses of Diagnostic Test Accuracy Studies

There are different methods for meta-analyses that have been proposed for diagnostic accuracy studies including the older traditional methods and more advanced methods (Macaskill et al., 2010; Ochodo et al., 2013; Takwoingi et al., 2015).

3.2.4.1 Traditional Methods

Traditional methods for meta-analysis generally include those based on a univariate approach and summary ROC curve based on simple linear regression (Macaskill et al., 2010, Ochodo et al., 2013, Takwoingi et al., 2015).

The univariate approach entails independent pooling of one outcome, for example, pooling of sensitivity and specificity separately, pooling of likelihood ratios and pooling of diagnostic odds ratios. Independent pooling ignores the threshold effect or correlation between sensitivity and specificity and may provide misleading results (Macaskill et al., 2010; Takwoingi et al., 2015). Thus, this method is generally not recommended unless in cases where bivariate models fail to converge and provide a model estimate. An example of this scenario is when studies with small sample sizes are included in the meta-analyses (Takwoingi et al., 2017). In addition, if included studies report no false positives (all report 100% specificity), a univariate random-effects model can be used to pool estimates of sensitivity only.

Generating a summary ROC curve based on linear regression (the Moses-Littenberg model) (Moses et al., 1993) accounts for variation in threshold but does not account for heterogeneity in test accuracy which may be caused by other factors such as population, spectrum of disease, conduct and interpretation of the tests (Macaskill et al., 2010; Takwoingi et al., 2015; Naaktgeboren et al., 2016).

3.2.4.2 Advanced Methods

The advanced methods generally take into account between- and within-study heterogeneity and the threshold effect between sensitivity and specificity. They are thus considered more robust than traditional methods (Macaskill et al., 2010). The advanced methods for meta-analysis include the bivariate random-effects model; hierarchical summary ROC curve (HSROC) model; the trivariate analysis of sensitivity, specificity and prevalence (Chu et al., 2009); and the multivariate random-effects model for tests with multiple thresholds or cut-offs (Hamza et al., 2009). The Cochrane methods group generally recommends the bivariate and HSROC models for meta-analyses of diagnostic test accuracy studies, the choice of which depends on the thresholds reported (Macaskill et al., 2010).

The bivariate model pools estimates of sensitivity and specificity directly around a common threshold. It also generates a confidence region and prediction region around the summary point (Reitsma et al., 2005; Macaskill et al., 2010). For example if most or all included studies evaluating the accuracy of point of care tests for schistosomiasis report (Ochodo et al., 2015) estimates of sensitivity and specificity at a threshold of +1, then the bivariate model can pool estimates at this threshold (Fig. 7). Another example of a common threshold is when all test results are binary, reported as either present or absent.

The HSROC method is recommended where included studies report sensitivity and specificity estimated at different thresholds (Fig. 7). In such a scenario, modelling a summary accuracy point will not produce clinically meaningful results. The HSROC model estimates the underlying ROC curve that displays how each study's sensitivity and specificity vary at different thresholds (Macaskill et al., 2010, Takwoingi et al., 2015).

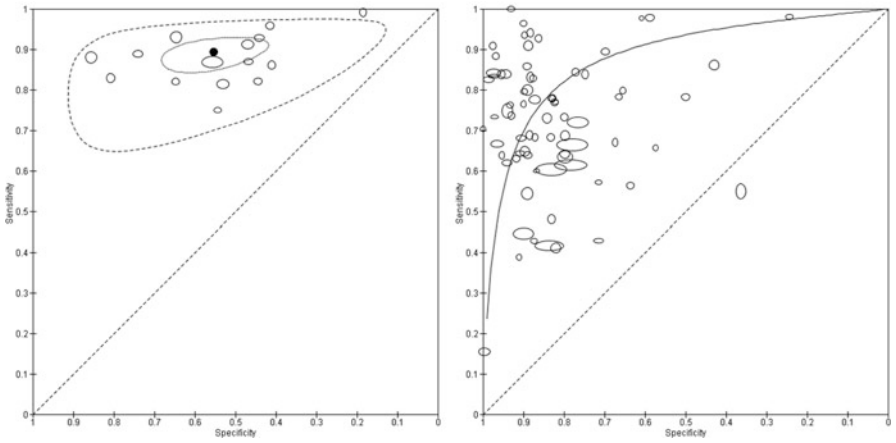


Fig. 7 Example of bivariate (left) and SROC plots (right). The thick black point in the left diagram shows the average value for sensitivity and specificity. The inner ellipse around the black spot represents the 95% confidence regions around the summary estimates. The outer ellipse represents the prediction region. The size of the points is proportional to the study sample size. The solid line in right shows the summary ROC curve. (Reproduced from Ochodo et al. (2015) under the terms of the Creative Commons Attribution Non-Commercial License CC BY-NC)

3.2.5 Steps in Conducting a Meta-Analyses of Diagnostic Test Accuracy Study

In this section we summarise the steps to be considered when conducting a meta-analysis of diagnostic test accuracy.

3.2.5.1 Analysis of a Single Test (Comparing an Index Test to a Reference Standard)

1. Identify and state the unit of analysis. Is it at individual participant level or at sample level? Note that in sample analysis, a participant can contribute more than one sample.
2. From each study identify the test positivity thresholds, and construct two-by-two tables at those thresholds.
3. Conduct preliminary exploratory analyses by plotting the estimates of sensitivity and specificity in forest plots and SROC plots. Note the difference in forest plots for diagnostic accuracy studies. Unlike those for intervention studies which present one measure of effect and summary effect measure, forest plots of diagnostic test accuracy plot two measures (sensitivity and specificity) and do not have a summary effect measure. Review Manager, a software tool by the

Cochrane collaboration, can be used to plot forest and SROC plots (Takwoingi et al., 2015).
4. Visually assess the forest plots to assess the extent of variability of estimates of sensitivity and specificity. If there isn't a lot of variability and if studies are sufficient (≥4 studies), a meta-analysis can be conducted.
5. Depending on the thresholds reported, select whether to use the bivariate method to pool estimates at a common threshold or the HSROC method to estimate accuracy at different thresholds. The statistical packages that can fit the bivariate model include Stata, SAS, R, WinBUGS and MLWin. Those that can fit the HSROC model include SAS, R and WinBUGS (Takwoingi et al., 2015).
6. If the advanced models fail to converge, a simple univariate model for sensitivity and specificity can be fitted separately using a random-effects meta-analysis.

3.2.5.2 Analysis of Two or More Tests (Comparing Index Tests)

The performance of accuracy of two or more tests can be compared directly where only studies that applied both index tests in the same individuals are used or compared indirectly where all studies are included in the analysis. These tests can be compared by adding the covariate test type to the bivariate or HSROC models (Macaskill et al., 2010).

3.2.5.3 Investigations of Heterogeneity

Source of heterogeneity can be investigated by adding the covariates to the bivariate or HSROC models by using subgroup or sensitivity analyses (Macaskill et al., 2010; Naaktgeboren et al., 2014).

3.2.6 Conclusion

Meta-analyses of diagnostic test accuracy studies mostly involve the pooling of two outcomes. Sensitivity and specificity are the most commonly reported diagnostic outcomes and hence used in meta-analysis. Meta-analysis can be used for the analysis of a single test, comparing multiple tests and investigating sources of heterogeneity. The decision to conduct a meta-analysis of diagnostic test accuracy studies depends on the size and number of included studies as well as the heterogeneity observed visually in the forest plots. The advanced methods for meta-analysis that account for test threshold effect as well as between- and within-study heterogeneity are generally recommended for diagnostic test accuracy studies.

3.3 Data Synthesis when Meta-Analysis Is Not Possible

Samuel A. Abariga

Meta-analysis has the advantage of mathematically synthesizing data across studies that contributed data for the outcome of interest. However, certain factors may render meta-analysis either inappropriate or impossible.

3.3.1 Reasons That May Render Meta-Analysis Impossible or Inappropriate

Broadly speaking, two reasons may explain why it is inappropriate to combine results of individual studies in a meta-analysis. These are (1) heterogeneity among included studies and (2) insufficient information on review specific outcomes of interest.

3.3.1.1 Substantial Heterogeneity Among Included Studies

Substantial diversity in important study characteristics or heterogeneity among included studies such as study design, study population, interventions, outcomes or study setting may be some of the concerns for which it may not be appropriate to conduct a meta-analysis (Ioannidis et al., 2008). Heterogeneity among studies may be grouped into four types, namely, clinical, geographic, methodologic or statistical heterogeneity (Herbert & Kari, 2005). Clinical heterogeneity may occur in a variety of ways, for instance, intervention studies conducted in different settings e.g. studies among participants in a community or a tertiary care facility may be different in terms of patient outcomes due to the differences in the level of care provided. Tertiary hospitals may have more experienced and specialised healthcare providers compared to community hospitals, and therefore the level of care may vary. Other differences in participant characteristics such as age, gender, race and difference in disease severity (e.g. cancer patients with stage 1 and stage 4 disease lumped together) may also be an important source of clinical diversity. Variability of frequency and route of administration of an intervention such as a drug, where participants in some of the included studies received the intervention via the oral route and others via the intravenous route or variability in dosing frequencies, may all be important sources of clinical heterogeneity. Outcomes of interest may also be a source of variability if the outcomes are defined differently in individual studies. For instance, in studies involving HIV patients where investigators are interested in the effect of an intervention (e.g. antiretroviral treatment) on the progression of HIV disease, if some studies use objective laboratory markers such as viral loads or CD4 counts to evaluate disease progression and other studies use resolution of clinical symptoms of disease as a proxy for disease progression, then clinical heterogeneity will be present among these studies when combined in a

meta-analysis. Important methodologic sources of heterogeneity may arise due to differences in study designs, e.g. RCTs, cross-sectional and prospective cohort study designs. Statistical heterogeneity is the percentage of variability across studies that is not due to chance. Geographical diversity among individual studies may arise when there is clinical practice variability among different regions. In the presence of these variabilities, as well as the presence of very high statistical heterogeneity (described elsewhere in this chapter), combining data from individual studies into a pooled effect estimate may be inappropriate.

3.3.1.2 Insufficient Information on Review Specific Outcomes of Interest

Among included studies in a systematic review, individual studies may have insufficient or incomplete information on treatment effects. For instance, studies may report the estimate without confidence intervals, or studies may simply report the magnitude and direction of the treatment effect along with only the P value, or there may be differences in the effect measure across the individual studies included in the systematic review. The latter may arise when inappropriate statistical methods are used to analyse data from studies reporting different measures of association such as odds ratios, relative risk, hazard ratios, etc.

3.3.2 Recommended Alternative Ways of Synthesizing Data

When meta-analysis is not possible for obvious reasons, in the interest of transparency, review authors should clearly state in the methods section of the review the methods intended to be used to present results. There are two main ways to present results without meta-analysis: vote counting based on direction of effect or visual display and presentation of the data (McKenzie & Brennan, 2019). Whichever method is used, it is recommended that review authors follow the reporting guidelines on synthesis without meta-analysis in a systematic way and clearly describe the following: methods used to group studies, standardised metric used for the synthesis, the synthesis method, method of data presentation, a summary of the synthesis findings as well as limitations of the synthesis (Campbell et al., 2020).

3.3.2.1 Vote Counting Based on Direction of Effect

When included studies report evidence of an effect and there is inconsistency in the effect measures across studies, vote counting based on direction of effect can be used to compare the number of effects showing benefit in favour of the intervention and the number that shows harm. The definition of 'benefit' or 'harm' may render this approach problematic (McKenzie & Brennan, 2019). Vote counting based on statistical significance or the size of the effect estimate is not recommended (McKenzie & Brennan, 2019).

3.3.2.2 Visual Display of Results

Visual display and presentation of results can be done by using either a structured tabulation of results across studies, forest plots, box-and-whiskers, bubble, albatross or harvest plots. Structured tabulation of studies can be done by ordering them by an important characteristic of the study such as the population, study design, intervention, outcomes, risk of bias or certainty of the evidence that is relevant for interpreting the findings. Forest plots can be used to display the results of individual studies without an overall pooled effect estimate (see image in other sections).

The box-and-whisker plots (Fig. 8a) depicts the distribution of effect estimates where the whiskers indicate extreme values, the line within the box the median and the upper and lower limits of the box indicating the 25th and 75th percentiles, respectively (McGill et al., 1978). Figure 8a displays results of state-specific age-adjusted odds ratios for social gradients (by wealth, education and social caste for diabetes, hypertension and obesity in India (Corsi & Subramanian, 2019).

The bubble plots (Fig. 8b) are similar to the box-and-whisker plots and are more suitable for visually displaying results when there are fewer studies. The bubble plot (Fig. 8b) (Higgins et al., 2019c) depicts multiple dimensions using size and colour

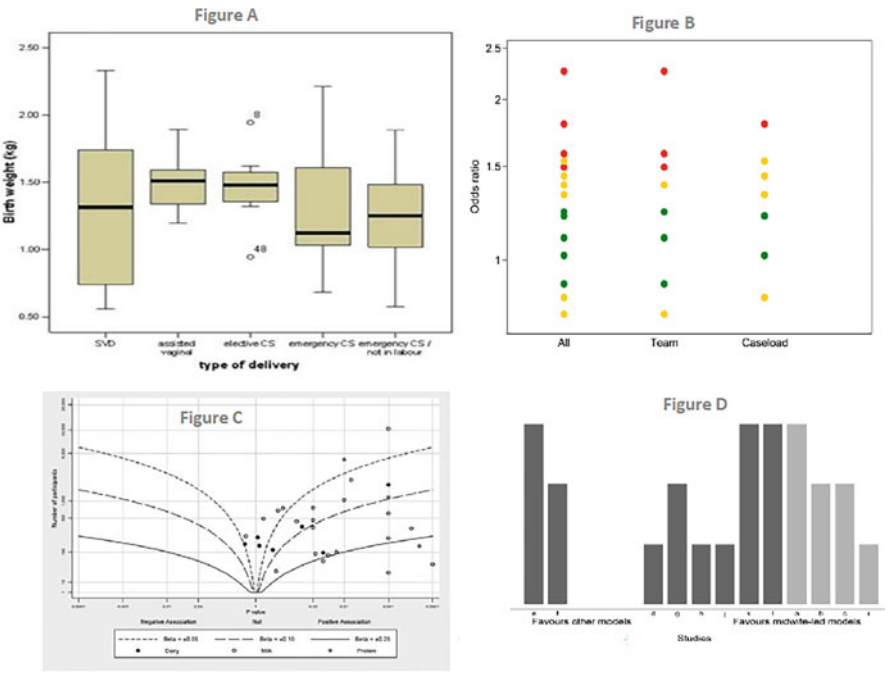


Fig. 8 (**a**–**d**) Examples of plots used to display results when meta-analysis is not feasible. (Reproduced from Corsi and Subramanian (2019) (**a**), Higgins et al. (2019c) (**b**, **d**) and Harrison et al. (2017) and (**c**) under the terms of the Creative Commons Attribution Non-Commercial License CC BY-NC)

of the bubbles. In Figure b, bubble sizes and colours reflect design: randomised trials (large, green), quasi-randomised trial (medium, red), non-randomised studies (small, blue); and precision (as well as multiple categories of the outcome such as potential harm, no effect, potential benefit and benefit) (Higgins et al., 2019c).

The albatross (Fig. 8c) and the harvest plots (Fig. 8d) are also alternative ways of visually displaying results when synthesizing results without meta-analysis and reflects the nature of the evidence. When included studies differ in terms of their designs, the heights of the bars can be used to show the differences in study designs, e.g. observational studies (short), quasi-randomised trials (medium) versus randomised controlled trials (tall). Where studies are similar in their design, e.g. RCTs, the heights of the bars can also be used to represent the sample size. The shades may be used to depict the duration of follow-up, e.g. longest follow-up (black) to shortest follow-up (light grey) or no information (white) (Ogilvie et al., 2008; Harrison et al., 2017). Figure 8d is an example of visual presentation of results using albatross plots. The graph depicts results from studies assessing the effect of exercise training on left ventricular fraction after acute myocardial infarction (Harrison et al., 2017). The harvest plot groups effect according to their direction (Fig. 8d).

4 Overview of Certainty of Evidence and GRADE

Michael McCaul

4.1 What Is GRADE and Why Is It Needed?

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) began in 2000 as an informal collaboration of people with an interest in addressing the limitations of evidence grading systems in healthcare. Today the GRADE approach is considered the standard in guideline development and rating the quality of evidence in evidence synthesis products and when making recommendations in guidelines.

Healthcare and public health judgements about evidence and recommendations are complex. Systematic reviews might be able to provide the complete picture of the current evidence for a particular public health question, however alone do not provide sufficient information for making well-informed decisions. Furthermore, systematic review authors and guideline developers are inconsistent in how they rate the quality of evidence (from systematic reviews) and how they grade the strength of recommendations in guidelines. This is partly due to the notion of quality or certainty of evidence not being clear, various heterogeneous evidence grading systems, lack of transparency in making judgements and tools that do not separate the certainty of the evidence from the strength of recommendations (Guyatt

Table 7 Definitions of GRADE evidence ratings

Grade	Definition
High (⊕⊕⊕⊕)	We are very confident that the true effect lies close to that of the estimate of the effect
Moderate (⊕⊕⊕⊖)	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low (⊕⊕⊖⊖)	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low (⊕⊖⊖⊖)	We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

et al., 2008). Towards solutions, the GRADE system, through a transparent process, classifies the certainty of evidence in a systematic review as one of four levels – high, moderate, low or very low categories (Guyatt et al., 2011b). These are defined in Table 7, as per the GRADE Handbook (Schünemann et al., 2013).

4.2 GRADE Judgements and Domains

When applying the GRADE approach, evidence from RCTs start as high certainty of evidence, but confidence in the evidence may be decreased (termed downgraded) for several reasons (Balshem et al., 2011). The domains that need to be considered are described below.

4.2.1 Domains for downgrading evidence

4.2.1.1 Risk of Bias

Limitations in study design and how the study is conducted may lead to biased estimates of the treatment effect. Our confidence in the treatment effect is reduced if there are major limitations in study design and execution. For RCTs, these include bias originating from improper randomisation to lack of allocation concealment and lack of blinding to incomplete outcome data or follow-up among others. In observational research, study limitations that decrease our confidence in effect estimates include flawed measurements of both exposure and outcomes, failure to control for important confounding or incomplete or differential follow-up. Depending on the extent of bias, the certainty of evidence can be downgraded by one or two levels (Guyatt et al., 2011h).

4.2.1.2 Inconsistency

Inconsistency refers to unexplained heterogeneity in results. Various measures used to quantify the extent of heterogeneity in a meta-analysis, were discussed in previous sections and should be considered when making a judgement around inconsistency. Importantly, reasons for inconsistency should be explored considering differences in the population, intervention, comparisons and outcomes (Guyatt et al., 2011a).

4.2.1.3 Indirectness

We can place more certainty in study results when we have direct evidence that speaks to our population, intervention, setting and outcomes of interest. Systematic review authors and guideline development teams should consider the extent to which evidence is applicable to their review question. If there is extensive indirectness, authors can downgrade by one or two levels (Guyatt et al., 2011e).

4.2.1.4 Imprecision

Imprecision results when studies include relatively few participants or few events and have a wide corresponding 95% confidence interval. Wide confidence intervals decrease our certainty in the results but should be interpreted together with considering the optimal information size (Guyatt et al., 2011d).

4.2.1.5 Reporting (Publication) Bias

Publication bias refers to the systematic over- or under-estimation of study results due to the selective reporting and publication of studies. Our confidence in the effect estimate in a systematic review is decreased if we suspect publication bias. Publication bias can be investigated using funnel plots (Guyatt et al., 2011f).

4.2.2 Domains for upgrading observational studies

Observational studies (such as cohort studies) start as low certainty of evidence but can be upgraded to a higher certainty of evidence by considering certain criteria such as a large magnitude of effect, dose response effects and the effect of plausible residual confounding (Guyatt et al., 2011g). Public health interventions are often investigated using observational research or quasi-experimental designs. In the case of quasi-experimental designs, these also provide high certainty of evidence as a starting point but will automatically be downgraded for limitations in design (risk of bias). Each domain can be downgraded by one or two levels; ideally this should be done by two reviewers independently and reach a consensus view on downgrading decisions.

Summary of findings:

Pre-hospital compared to in-hospital thrombolysis mortality for ST-elevation myocardial infarction

Patient or population: ST-elevation myocardial infarction

Intervention: Pre-hospital

Comparison: in-hospital thrombolysis mortality

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	Nº of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with in-hospital thrombolysis mortality	Risk with Pre-hospital				
All cause hospital mortality follow up: range 1 days to 30 days	73 per 1,000	**53 per 1,000** (27 to 103)	**RR 0.73** (0.37 to 1.41)	538 (3 RCTs)	⨁⨁◯◯ LOW [a,b]	Prehospital thrombolysis may result in a large reduction in all cause hospital mortality.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

a. Downgraded by 1 level for risk of bias due to poor reporting of random sequence generation, allocation concealment (not described and concealment broken) and inadequate outcome reporting

b. Downgraded by 1 level for imprecision as CI includes appreciable benefit and appreciable harm.

Fig. 9 Example of GRADE SoF table. (Data taken from McCaul et al. (2014))

4.2.3 Making and presenting GRADE judgements

Importantly, GRADE certainty of evidence judgements are made across important outcomes (that are deemed critical or important) (Guyatt et al., 2011c), and any reasons for downgrading or upgrading are transparently reported. For example, the quality of evidence would be graded 'high' if the evidence was from several RCTs with low risk of bias; however the certainty would become lower if there were concerns regarding inconsistency of results, indirectness, imprecision or publication bias. In systematic reviews, these transparent GRADE judgements are presented as 'Summary of Findings' (SoF) tables, for both binary (Guyatt et al., 2013a) and continuous outcomes (Guyatt et al., 2013b).

Figure 9 presents a GRADE SoF table for a systematic review of pre-hospital versus in-hospital thrombolytic agents for mortality in myocardial infarct (heart attack) patients, with corresponding meta-analysis, GRADE certainty of evidence and interpretation (McCaul et al., 2014).

These summaries present the main findings of a review in a transparent, structured and simple manner. They provide key information per outcome regarding the certainty of evidence (quality), the magnitude of effect and comparisons of characteristics and details linked to the PICO: question (Guyatt et al., 2011c). A SoF can be presented in various ways and can be supported by more detailed tables, known as 'Evidence Profiles', which provide more detailed explanations following the GRADE domains. Evidence profiles are more commonly used for guideline development teams when making recommendations via the linked GRADE Evidence to Decision (EtD) framework using software such as GRADEPro GDT (Andrews et al., 2013a, b; Alonso-Coello et al., 2018). Software to produce SoF and evidence tables can be found in here (www.gradepro.org), including GRADE's interactive handbook providing detail around how to make GRADE judgements and navigate the GRADE domains when judging quality of evidence (Broazek et al., 2011).

Once an outcome's quality of evidence has been GRADEd, informative statements should be generated to communicate the findings of the systematic review. For intervention reviews, the GRADE working group has produced useful guidance on how to do this (Santesso et al., 2020). These statements incorporate both the certainty of the evidence and the size of the effect estimate in creating suggested statements.

GRADE guidance has been developed for systematic reviews addressing various types of questions. Detailed descriptions of GRADE for authors of guidelines and systematic reviews have been published in the Journal of Clinical Epidemiology ('GRADE Series', 2021). Useful GRADE guidance has been provided for diagnostic accuracy studies (Brozek et al., 2009; Schünemann et al., 2020), network meta-analysis (Brignardello-Petersen et al., 2018) and qualitative evidence synthesis (Lewin et al., 2015; Noyes et al., 2018).

5 Useful Resources

Below is a list of useful resources related to the content of this chapter.

Name	Description	URL
Cochrane Library	Collection of databases containing different types of evidence to inform healthcare decision-making, including the Cochrane Database of Systematic reviews and Cochrane Central Register of Controlled Trials	https://www.cochranelibrary.com

(continued)

(continued)

Name	Description	URL
Epistemonikos	Database of systematic reviews relevant for health decision-making that is regularly updated	https://www.epistemonikos.org
Health Evidence™	Database of pre-appraised systematic reviews on interventions relevant to public health	https://www.healthevidence.org
Cochrane training	Webinars and resources linked to conducting systematic reviews	https://training.cochrane.org
Cochrane Handbook	Detailed description and guidance on conducting systematic reviews and performing meta-analysis	https://training.cochrane.org/handbook
PRISMA statement	Reporting guideline for systematic reviews and meta-analysis	http://www.prisma-statement.org/
Cochrane Handbook for Diagnostic accuracy reviews	Detailed description and guidance on conducting systematic reviews and performing meta-analysis of diagnostic test accuracy studies	https://methods.cochrane.org/sdt/handbook-dta-reviews
PRISMA for Diagnostic Test Accuracy	Reporting guideline for systematic reviews and meta-analysis of diagnostic test accuracy studies	http://www.prisma-statement.org/Extensions/DTA
CINeMA	Confidence in Network Meta-Analysis web application	https://cinema.ispm.unibe.ch/
GRADEpro GDT	GRADE web application for SRs and guidelines	https://gradepro.org/
MetaInsight	Free NMA tool, including DTA-MA	https://crsu.shinyapps.io/metainsightc/
NMA Toolkit	Cochrane Network Meta-analysis Toolkit	https://methods.cochrane.org/cmi/network-meta-analysis-toolkit
STATA routines for NMA	Multiple treatments Meta-analysis in STATA	https://mtm.uoi.gr/index.php/stata-routines-for-network-meta-analysis

References

Alonso-Coello, P., Schünemann, H. J., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., Treweek, S., Mustafa, R. A., Rada, G., Rosenbaum, S., Morelli, A., Guyatt, G. H., & Oxman, A. D. (2018). GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *Gaceta Sanitaria, 32*(2), 166.e161–166.e110.

Andrews, J., Guyatt, G., Oxman, A. D., Alderson, P., Dahm, P., Falck-ytter, Y., Nasser, M., Meerpohl, J., Post, P. N., & Kunz, R. (2013a). GRADE SERIES GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *Journal of Clinical Epidemiology, 66*, 719–725.

Andrews, J. C., Sch, H. J., Oxman, A. D., Pottie, K., Meerpohl, J. J., Alonso, P., Rind, D., Montori, V. M., Pablo, J., Norris, S., Elbarbary, M., Post, P., Nasser, M., Shukla, V., Jaeschke, R., Brozek, J., Djulbegovic, B., & Guyatt, G. (2013b). GRADE guidelines: 15. Going from evidence to recommendation d determinants of a recommendation ' s direction and strength. *Journal of Clinical Epidemiology, 66*, 726–735.

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology, 64*(4), 401–406.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111.

Brignardello-Petersen, R., Bonner, A., Alexander, P. E., Siemieniuk, R. A., Furukawa, T. A., Rochwerg, B., Hazlewood, G. S., Alhazzani, W., Mustafa, R. A., Murad, M. H., Puhan, M. A., Schünemann, H. J., & Guyatt, G. H. (2018). Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *Journal of Clinical Epidemiology, 93*, 36–44.

Broazek, J. L., Akl, E. A., Compalati, E., Kreis, J., Terracciano, L., Fiocchi, A., Ueffing, E., Andrews, J., Alonso-Coello, P., Meerpohl, J. J., Lang, D. M., Jaeschke, R., Williams, J. W., Phillips, B., Lethaby, A., Bossuyt, P., Glasziou, P., Helfand, M., Watine, J., Afilalo, M., Welch, V., Montedori, A., Abraha, I., Horvath, A. R., Bousquet, J., Guyatt, G. H., & Schünemann, H. J. (2011). Grading quality of evidence and strength of recommendations in clinical practice guidelines part 3 of 3. The GRADE approach to developing recommendations. *Allergy: European Journal of Allergy and Clinical Immunology, 66*(5), 588–595.

Brownson, R. C., Gurney, J. G., & Land, G. H. (1999). Evidence-based decision making in public health. *Journal of Public Health Management and Practice, 5*(5), 86–97.

Brozek, J. L., Akl, E. A., Jaeschke, R., Lang, D. M., Bossuyt, P., Glasziou, P., Helfand, M., Ueffing, E., Alonso-Coello, P., Meerpohl, J., Phillips, B., Horvath, A. R., Bousquet, J., Guyatt, G. H., & Schünemann, H. J. (2009). Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy: European Journal of Allergy and Clinical Immunology, 64*(8), 1109–1116.

Bucher, H. C., Guyatt, G. H., Griffith, L. E., & Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology, 50*(6), 683–691.

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., & Thomas, J. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ, 368*.

Chaimani, A., Caldwell, D. M., Li, T., Higgins, J. P. T., & Salanti, G. (2019). Chapter 11: Undertaking network meta-analyses. *Cochrane.*

Chinnock, P., Siegfried, N., & Clarke, M. (2005). Is evidence-based medicine relevant to the developing world? *PLoS Medicine, 2*(5), e107.

Chu, H., Nie, L., Cole, S. R., & Poole, C. (2009). Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection. *Statistics in Medicine, 28*(18), 2384–2399.

Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E., & Welton, N. J. (2009). Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine, 28*(March), 1861–1881.

Corsi, D. J., & Subramanian, S. (2019). Socioeconomic gradients and distribution of diabetes, hypertension, and obesity in India. *JAMA Network Open, 2*(4), e190411–e190411.

Dawes, M., Summerskill, W., Glasziou, P., Cartabellotta, A., Martin, J., Hopayian, K., Porzsolt, F., Burls, A., Osborne, J., & T. Second International Conference of Evidence-Based Health Care and Developers. (2005). Sicily statement on evidence-based practice. *BMC Medical Education, 5*(1), 1.

De Buck, E., Van Remoortel, H., Hannes, K., Govender, T., Naidoo, S., Avau, B., Veegaete, A. V., Musekiwa, A., Lutje, V., Cargo, M., Mosler, H.-J., Vandekerckhove, P., & Young, T. (2017). Approaches to promote handwashing and sanitation behaviour change in low- and middle-income countries: A mixed method systematic review. *Campbell Systematic Reviews, 13*(1), 1–447.

Deeks, J., Higgins, J., & Altman, D. (2008). Analysing data and undertaking meta-analyses, chapter 9. *Cochrane handbook for systematic reviews of interventions version, 510*.

Deeks, J., Higgins, J., & Altman, D. (2019). Chapter 10: Analysing data and undertaking meta-analyses. In H. JPT, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.0 (updated July 2019)*. Cochrane.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.

Dickersin, K., & Berlin, J. A. (1992). Meta-analysis: State-of-the-science. *Epidemiologic Reviews, 14*(1), 154–176.

Donegan, S., Williamson, P., D'Alessandro, U., & Tudur Smith, C. (2013). Assessing key assumptions of network meta-analysis: A review of methods. *Research Synthesis Methods, 4*(4), 291–323.

ECDPC. (2011). *Evidence-based methodologies for public health – How to assess the best available evidence when time is limited and there is lack of sound evidence*. Stockholm.

Egger, M., Davey-Smith, G., & Altman, D. (2008). *Systematic reviews in health care: Meta-analysis in context*. Wiley.

Elliott, W. J., & Meyer, P. M. (2007). Incident diabetes in clinical trials of antihypertensive drugs: A network meta-analysis. *The Lancet, 369*(9557), 201–207.

Gadkary, C., Alderson, P., & Signorini, D. F. (2002). Therapeutic hypothermia for head injury. *Cochrane Database of Systematic Reviews, 1*.

GRADE Series (2021). From https://www.jclinepi.com/content/jce-GRADE-Series

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck, Y., Alonso-coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal, 336*(April), 924–926.

Guyatt, G., Oxman, A., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-coello, P., Glasziou, P., Jaeschke, R., Akl, E. A., Norris, S., Vist, G., Dahm, P., Shukla, V. K., & Higgins, J. (2011a). GRADE guidelines: 7. Rating the quality of evidence d inconsistency. *Journal of Clinical Epidemiology, 64*, 1294–1302.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., Debeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011b). GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383–394.

Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., Vist, G., Alderson, P., Glasziou, P., Falck-Ytter, Y., & Schünemann, H. J. (2011c). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology, 64*(4), 395–400.

Guyatt, G. H., Oxman, A. D., Kunz, R., Brozek, J., Alonso-Coello, P., Rind, D., Devereaux, P. J., Montori, V. M., Freyschuss, B., Vist, G., Jaeschke, R., Williams, J. W., Murad, M. H., Sinclair, D., Falck-Ytter, Y., Meerpohl, J., Whittington, C., Thorlund, K., Andrews, J., & Schünemann, H. J. (2011d). GRADE guidelines 6. Rating the quality of evidence – imprecision. *Journal of Clinical Epidemiology, 64*(12), 1283–1293.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Falck-Ytter, Y., Jaeschke, R., Vist, G., Akl, E. A., Post, P. N., Norris, S., Meerpohl, J., Shukla, V. K., Nasser, M., & Schünemann, H. J. (2011e). GRADE guidelines: 8. Rating the quality of evidence – indirectness. *Journal of Clinical Epidemiology, 64*(12), 1303–1310.

Guyatt, G. H., Oxman, A. D., Montori, V., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., Djulbegovic, B., Atkins, D., Falck-Ytter, Y., Williams, J. W., Meerpohl, J., Norris, S. L., Akl, E. A., & Schünemann, H. J. (2011f). GRADE guidelines: 5. Rating the quality of evidence – publication bias. *Journal of Clinical Epidemiology, 64*(12), 1277–1282.

Guyatt, G. H., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., Atkins, D., Kunz, R., Brozek, J., Montori, V., Jaeschke, R., Rind, D., Dahm, P., Meerpohl, J., Vist, G., Berliner, E., Norris, S., Falck-Ytter, Y., Murad, M. H., & Schünemann, H. J. (2011g). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology, 64*(12), 1311–1316.

Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., Montori, V., Akl, E. A., Djulbegovic, B., Falck-Ytter, Y., Norris, S. L., Williams, J. W., Atkins, D., Meerpohl, J., & Schünemann, H. J. (2011h). GRADE guidelines: 4. Rating the quality of evidence – study limitations (risk of bias). *Journal of Clinical Epidemiology, 64*(4), 407–415.

Guyatt, G. H., Oxman, A. D., Santesso, N., Helfand, M., Vist, G., Kunz, R., Brozek, J., Norris, S., Meerpohl, J., Djulbegovic, B., Alonso-Coello, P., Post, P. N., Busse, J. W., Glasziou, P., Christensen, R., & Schünemann, H. J. (2013a). GRADE guidelines: 12. Preparing summary of findings tables – binary outcomes. *Journal of Clinical Epidemiology, 66*(2), 158–172.

Guyatt, G. H., Thorlund, K., Oxman, A. D., Walter, S. D., Patrick, D., Furukawa, T. A., Johnston, B. C., Karanicolas, P., Akl, E. A., Vist, G., Kunz, R., Brozek, J., Kupper, L. L., Martin, S. L., Meerpohl, J. J., Alonso-coello, P., Christensen, R., & Schunemann, H. J. (2013b). GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles- continuous outcomes. *Journal of Clinical Epidemiology, 66*(2), 173–183.

Hamza, T. H., Arends, L. R., van Houwelingen, H. C., & Stijnen, T. (2009). Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology, 9*, 73.

Harrison, S., Jones, H. E., Martin, R. M., Lewis, S. J., & Higgins, J. P. (2017). The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. *Research Synthesis Methods, 8*(3), 281–289.

Herbert, R. D., & Kari, B. (2005). Analysis of quality of interventions in systematic reviews. *BMJ, 331*(7515), 507–509.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560.

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172*(1), 137–159.

Higgins, J. P. T., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E., & White, I. R. (2012). Consistency and inconsistency in network meta-analysis: Concepts and models for multi-arm studies. *Research Synthesis Methods, 3*(2), 98–110.

Higgins, J., Eldridge, S., & T. Li. (2019a). Chapter 23: Including variants on randomized trials. *Cochrane Handbook for Systematic Reviews of Interventions version 6.0* (updated July 2019).

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2019b). *Cochrane handbook for systematic reviews of interventions version 6.0* (updated July 2019), Cochrane.

Higgins, J. P., López-López, J. A., Becker, B. J., Davies, S. R., Dawson, S., Grimshaw, J. M., McGuinness, L. A., Moore, T. H., Rehfuess, E. A., & Thomas, J. (2019c). Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Global Health, 4*(Suppl 1), e000858.

Ioannidis, J. P., & Lau, J. (1999). Pooling research results: Benefits and limitations of meta-analysis. *Joint Commission Journal on Quality Improvement, 25*(9), 462–469.

Ioannidis, J. P., Patsopoulos, N. A., & Rothstein, H. R. (2008). Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ, 336*(7658), 1413–1415.

Islam, M. Z., Musekiwa, A., Islam, K., Ahmed, S., Chowdhury, S., Ahad, A., & Biswas, P. K. (2014). Regional variation in the prevalence of E. coli O157 in cattle: a meta-analysis and meta-regression. *PLoS ONE, 9*(4), e93299.

Jackson, D., & White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine, 20*(January), 24812498–24812498.

Jenicek, M. (1997). Epidemiology, evidenced-based medicine, and evidence-based public health. *Journal of Epidemiology, 7*(4), 187–197.

Leeflang, M., Deeks, J., Gatsonis, C., Bossuyt, P., & C. D. T. A. W. Group. (2008). Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine, 149*(12), 889–897.

Lengeler, C. (2004). Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database of Systematic Reviews, 2*.

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C. J., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R., & Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: An approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Medicine, 12*(10), e1001895–e1001895.

Li, T., Puhan, M. A., Vedula, S. S., Singh, S., Dickersin, K., & T. A. H. N. M.-a. M. M. W. Group. (2011). Network meta-analysis-highly attractive but more methodological research is needed. *BMC Medicine, 9*(1), 79–79.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Chapter 10: Analysing and presenting results; *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0*. The Cochrane Collaboration

Mavridis, D., & Salanti, G. (2011). A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research, 22*(1), 133–158.

Mavridis, D., Giannatsi, M., Cipriani, A., & Salanti, G. (2015). A primer on network meta-analysis with emphasis on mental health. *Evidence-Based Mental Health, 18*(2), 40–46.

McCaul, M., Lourens, A., & Kredo, T. (2014). Pre-hospital versus in-hospital thrombolysis for ST-elevation myocardial infarction. *Cochrane Database of Systematic Reviews, 9*.

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician, 32*(1), 12–16.

McKenzie, J. and S. Brennan (2019). Chapter 12: Synthesizing and presenting findings using other methods. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019).

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Quality of reporting of meta-analyses. *Lancet, 354*(9193), 1896–1900.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & P. Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006–1012.

Moses, L., Shapiro, D., & Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine, 12*(14), 1293–1316.

Mudie, K., Jin, M. M., Tan, L., Kendall, J., Addo, I., Dos-Santos-Silva, J., Quint, L., Smeeth, S., Cook, D., Nitsch, B., Natamba, F., Gomez-Olive, X., Ako, A., & Perel, P. (2019). Non-communicable diseases in sub-Saharan Africa: A scoping review of large cohort studies. *Journal of Global Health, 9*(2), 020409.

Murdoch, I. E., Morris, S. S., & Cousens, S. N. (1998). People and eyes: Statistical approaches in ophthalmology. *British Journal of Ophthalmology, 82*(8), 971–973.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford University Press.

Naaktgeboren, C. A., van Enst, W. A., Ochodo, E. A., de Groot, J. A., Hooft, L., Leeflang, M. M., Bossuyt, P. M., Moons, K. G., & Reitsma, J. B. (2014). Systematic overview finds variation in approaches to investigating and reporting on sources of heterogeneity in systematic reviews of diagnostic studies. *Journal of Clinical Epidemiology, 67*(11), 1200–1209.

Naaktgeboren, C. A., Ochodo, E. A., Van Enst, W. A., de Groot, J. A. H., Hooft, L., Leeflang, M. M. G., Bossuyt, P. M., Moons, K. G. M., & Reitsma, J. B. (2016). Assessing variability in results in systematic reviews of diagnostic studies. *BMC Medical Research Methodology, 16*, 6.

Noyes, J., Booth, A., Cargo, M., Flemming, K., Garside, R., Hannes, K., Harden, A., Harris, J., Lewin, S., Pantoja, T., & Thomas, J. (2018). Cochrane qualitative and implementation methods group guidance serier – paper 1: Introduction. *Journal of Clinical Epidemiology, 97*, 35–38.

Ochodo, E. A., & Leeflang, M. M. G. (2012). Systematic reviews of diagnostic test accuracy for evidence-based diagnostic practice in Africa. *African Journal of Laboratory Medicine, 1*(1), 7.

Ochodo, E. A., Reitsma, J. B., Bossuyt, P. M., & Leeflang, M. M. (2013). Survey revealed a lack of clarity about recommended methods for meta-analysis of diagnostic accuracy data. *Journal of Clinical Epidemiology, 66*(11), 1281–1288.

Ochodo, E., Gopalakrishna, G., Spek, B., Reitsma, J., van Lieshout, L., Polman, K., Lamberton, P., Bossuyt, P., & Leeflang, M. (2015). Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas. *Cochrane Database of Systematic Reviews, 3*, CD009579.

Ogilvie, D., Fayter, D., Petticrew, M., Sowden, A., Thomas, S., Whitehead, M., & Worthy, G. (2008). The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology, 8*(1), 8.

Petropoulou, M., Nikolakopoulou, A., Veroniki, A. A., Rios, P., Vafaei, A., Zarin, W., Giannatsi, M., Sullivan, S., Tricco, A. C., Chaimani, A., Egger, M., & Salanti, G. (2017). Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *Journal of Clinical Epidemiology, 82*, 20–28.

Pryce, J., Richardson, M., & Lengeler, C. (2018). Insecticide-treated nets for preventing malaria. *Cochrane Database of Systematic Reviews, 11*.

Rao, J., & Scott, A. (1992). A simple method for the analysis of clustered binary data. *Biometrics*, 577–585.

Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., & Zwinderman, A. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology, 58*(10), 982–990.

Review Manager (RevMan). (2020). *The Cochrane Collaboration*.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1987). Meta-analyses of randomized controlled trials. *The New England Journal of Medicine, 316*(8), 450–455.

Salanti, G. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research, 17*, 279–301.

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods, 3*(2), 80–97.

Santesso, N., Glenton, C., Dahm, P., Garner, P., Akl, E. A., Alper, B., Brignardello-Petersen, R., Carrasco-Labra, A., De Beer, H., Hultcrantz, M., Kuijpers, T., Meerpohl, J., Morgan, R., Mustafa, R., Skoetz, N., Sultan, S., Wiysonge, C., Guyatt, G., & Schünemann, H. J. (2020). GRADE guidelines 26: Informative statements to communicate the findings of systematic reviews of interventions. *Journal of Clinical Epidemiology, 119*, 126–135.

Schünemann, H. J., Brožek, J., Guyatt, G. H., & A. Oxman. (2013). *GRADE Handbook*.

Schünemann, H. J., Mustafa, R. A., Brozek, J., Steingart, K. R., Leeflang, M., Murad, M. H., Bossuyt, P., Glasziou, P., Jaeschke, R., Lange, S., Meerpohl, J., Langendam, M., Hultcrantz,

M., Vist, G. E., Akl, E. A., Helfand, M., Santesso, N., Hooft, L., Scholten, R., Rosen, M., Rutjes, A., Crowther, M., Muti, P., Raatz, H., Ansari, M. T., Williams, J., Kunz, R., Harris, J., Rodriguez, I. A., Kohli, M., & Guyatt, G. H. (2020). GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *Journal of Clinical Epidemiology, 122*, 129–141.

Shah, M., Hanrahan, C., Wang, Z. Y., Dendukuri, N., Lawn, S. D., Denkinger, C. M., & Steingart, K. R. (2016). Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database of Systematic Reviews, 5*, CD011420.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ, 358*, j4008.

Šimundić, A. (2009). Measures of diagnostic accuracy: Basic definitions. *EJIFCC, 19*(4), 203–211.

Smithuis, F. M., Kyaw, M. K., Phe, U. O., van der Broek, I., Katterman, N., Rogers, C., Almeida, P., Kager, P. A., Stepniewska, K., Lubell, Y., Simpson, J. A., & White, N. J. (2013). The effect of insecticide-treated bed nets on the incidence and prevalence of malaria in children in an area of unstable seasonal transmission in western Myanmar. *Malaria Journal, 12*(1), 363.

Sobieraj, D. M., J. C. Cappelleri, W. L. Baker, O. J. Phung, C. M. White and C. I. Coleman (2013) Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. *Bmj Open 3*(7)

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. *Health Information and Libraries Journal, 36*(3), 202–222.

Takwoingi, Y., Riley, R., & Deeks, J. (2015). Meta-analysis of diagnostic accuracy studies in mental health. *Evidence-Based Mental Health, 18*(4), 103–109.

Takwoingi, Y., Guo, B., Riley, R. D., & Deeks, J. J. (2017). Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research, 26*(4), 1896–1911.

Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., Perrier, L., Hutton, B., Moher, D., & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine, 13*, 224.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Kastner, M., Levac, D., Ng, C., Sharpe, J. P., Wilson, K., Kenny, M., Warren, R., Wilson, C., Stelfox, H. T., & Straus, S. E. (2016). A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology, 16*, 15.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Lewin, S., Godfrey, C. M., Macdonald, M. T., Langlois, E. V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tuncalp, O., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473.

van Stralen, K. J., Stel, V. S., Reitsma, J. B., Dekker, F. W., Zoccali, C., & Jager, K. J. (2009). Diagnostic methods I: Sensitivity, specificity, and other measures of accuracy. *Kidney International, 75*(12), 1257–1263.

Venekamp, R. P., Sanders, S. L., Glasziou, P. P., Del Mar, C. B., & Rovers, M. M. (2015). Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev, 2015*(6), Cd000219.

Veroniki, A. A., Mavridis, D., Higgins, J. P. T., & Salanti, G. (2014). Characteristics of a loop of evidence that affect detection and estimation of inconsistency: A simulation study. *BMC Medical Research Methodology, 14*(1), 1–12.

White, I. R. (2011). Multivariate random-effects meta-regression: Updates to mvmeta. *Stata Journal, 11*(2), 255–270.

White, I. R., Barrett, J. K., & Higgins, J. P. T. (2012). Consistency and inconsistency in network meta-analysis: Model estimation using multivariate. *Research Synthesis Methods, 3*(June 2011), 111–125.

Whiting, P., Rutjes, A., Reitsma, J., Glas, A., Bossuyt, P., & Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Annals of Internal Medicine, 140*(3), 189–202.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., & Q.-S. Group. (2013). A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *Journal of Clinical Epidemiology, 66*(10), 1093–1104.

Whiting, P., Savovic, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., Churchill, R., & R. Group. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology, 69*, 225–234.

Whiting-O'Keefe, Q., Henke, C., & Simborg, D. W. (1984). Choosing the correct unit of analysis in medical care experiments. *Medical Care*, 1101–1114.

WHO. (2015). *The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. Policy guidance*. World Health Organization.

WHO. (2019a). *Guidelines for malaria vector control*. Licence: CC BY-NC-SA 3.0 IGO.

WHO. (2019b). *World malaria report*. Geneva.

WHO. (2020a). *WHO consolidated guidelines on tuberculosis* (Module 3: Diagnosis – Rapid diagnostics for tuberculosis detection). World Health Organization.

WHO. (2020b). *WHO operational handbook on tuberculosis* (Module 3: Diagnosis – rapid diagnostics for tuberculosis detection). World Health Organization.

Williams, T., McCaul, M., Schwarzer, G., Cipriani, A., Stein, D. J., & Ipser, J. (2020). Pharmacological treatments for social anxiety disorder in adults: A systematic review and network meta-analysis. *Acta Neuropsychiatr, 32*(4), 169–176.

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., Gluud, C., Martin, R. M., Wood, A. J. G., & Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *BMJ, 336*(7644), 601–605.

Young, T., Garner, P., Clarke, M., & Volmink, J. (2017). Series: Clinical epidemiology in South Africa. Paper 1: Evidence-based health care and policy in Africa: Past, present, and future. *Journal of Clinical Epidemiology, 83*, 24–30.

Zhen, J., Chan, C., Schoonees, A., Apatu, E., Thabane, L., & Young, T. (2020). Transmission of respiratory viruses when using public ground transport: A rapid review to inform public health recommendations during the COVID-19 pandemic. *South African Medical Journal, 110*(6), 478–483.

Statistical Meta-Analysis and Its Efficiency: A Real Data Analysis and a Monte-Carlo Simulation Study

Ding-Geng (Din) Chen

Abstract In the big-data era, the meta-data collected to address the same/similar scientific question usually come from diverse sources (such as, multi-regional clinical trials, multiple intervention studies). Meta-analysis (MA) is then a statistical methodology for combining information from these diverse sources to reach a more reliable conclusion. In this chapter, an overview of MA is given with emphasis on classical fixed-effects and random-effects MA models to synthesize summary statistics from all studies as well as meta-regression to explain the between-study heterogeneity. A Monte-Carlo simulation study is designed to illustrate the relative efficiency of the MA using summary statistics to the MA using the original individual participant-level data. Real meta-data from 13 clinical trials to assess Bacillus Calmette-Guerin vaccine in the prevention of tuberculosis are used to demonstrate the implementation of these meta-analysis models in open source R software.

Keywords Meta-analysis · Fixed-effects model · Random-effects model · Weighted-mean estimator · Dersimonian-Laird estimator · Heterogeneity · Meta-regression · Individual patient-level data · Monte-Carlo simulations

1 Introduction

In the big-data era, the meta-data collected to address the same/similar scientific question come from diverse sources. For example, in multi-regional clinical trials (MRCT), globalizing clinical trials tend to be larger in total sample size and targeted for multiple international regions for approval. As a result, they tend to be multi-regional and may cover the USA, Europe, China, Japan, other Asian-Pacific countries, South America, or African countries. Different countries and

D.-G. (Din) Chen (✉)
Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: din.chen@up.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_6

regions have different medical practice and regulatory requirements. Therefore, MRCT will have to be harmonized, which prompts actions from the FDA, EMA and Japanese Pharmaceuticals and Medical Devices Agency (PMDA) to release guidelines to synthesize and meta-analyze the data from these MRCT in an international harmonization.

Despite the globalization and harmonization of MRCT, there are emerging challenges in the conduct of such trials at different regions regarding potential regional and between-study heterogeneity. This heterogeneity, which needs to be carefully considered, is generated by different requirements from different regulatory agencies in different countries with regard to the study design and conduct of MRCT. Specifically, the heterogeneity can be generated from (1) different endpoints required by different regions since different regulatory authorities have different standards for primary, co-primary, or key secondary endpoints; (2) different time points required for hypothesis testing since the expected treatment duration or the primary time point for measuring treatment response in a trial may not be consistent across regions; (3) different experimental designs which are modified to match the different characteristics from different regions in MRCT; (4) different non-inferiority margins required by different regulatory agencies; and (5) different analytic patient populations defined in analysis plan by different regions. Any of these differences could lead to inflated type-I and type-II errors from the perspectives of statistical analysis and final recommendations. It then becomes imperative to enforce and follow the guidance from global harmonization from different health and regulatory authorities as in International Conference on Harmonization (http://www.ich.org).

Among all the methods to synthesize data from MRCT, meta-analysis (MA) is one of the statistical methods which is becoming more popular and extensively used for examining the validity of the effect-sizes from each trials, for quantifying the heterogeneity between the effect-sizes, and at the end, for providing an estimate of the overall pooled effect-size with optimal precision. With rising cost and time in developing MRCT, many trials are carried out with relative small numbers of patients resulting in low statistical power for detecting useful clinical trial effect-size. This phenomenon has increased the chance of producing conflicting results from different clinical trials. By pooling the estimated effect-sizes from each of the trials through meta-analytic methods, information from larger number of patients can be used and increased statistical power is expected.

In the literature, meta-analysis is traditionally referred to the meta-analysis of summary statistics from published studies (hereafter referred as "SS-MA"). The SS-MA is to combine point estimates of study effect-size from individual studies using the well-known fixed-effects and random-effects models which differ fundamentally whether to incorporate the between-study heterogeneity. This traditional meta-analysis has played an increasingly important role in biopharmaceutical, health, and medical sciences. Its applications have led to numerous scientific discoveries. For example, meta-analysis results are reported in more than five hundred articles in the *New England Journal of Medicine* in the past decade and many books are written to

SS-MA as seen in Whitehead, 2003; Hartung et al., 2008; Borenstein et al., 2009; Pigott, 2012; Chen and Peace, 2013.

From another standpoint, as data sharing and global harmonization become an increasingly common practice as discussed in MRCT, original individual patient-level data (IPD) can become more accessible. When IPD are available from all studies, it is intuitively preferable to perform meta-analysis using these IPD (hereafter referred as "IPD-MA") with some advanced hierarchical and multi-level statistical models. The IPD-MA is the gold standard in statistical modelling and analysis as pointed by Sutton and Higgins (2008) since all available data are used. Even though the IPD-MA is the gold standard in meta-analysis, to obtain or retrieve of the original IPD is costly, time-consuming, and impossible in most of the times which makes the IPD-MA impractical.

A logical, but fundamental question in meta-analysis is then how more efficient this IPD-MA relative to SS-MA. If they are relatively equivalent in statistical efficiency, we would resort to the SS-MA since the majority of meta-analyses are performed using summary statistics which are much easy to obtain. To understand this fundamental question, a series of research have been performed. Theoretically, Olkin and Sampson (1998) proved this equivalency for analysis of variance when there is no study-by-treatment interaction and the error variances are constant across studies. Mathew and Nordstrom (1999) immediately extended this homogeneous variance condition to heterogeneous variances from different studies. Simmonds and Higgins (2007) further examined a special case of linear regression models for continuous responses and Lin and Zeng (2010) then systematically investigated this fundamental question under a general likelihood inference setting for fixed-effects MA. Their framework was further extended by Liu et al. (2015) to a more complex setting of analyzing heterogenous studies and achieving complex evidence synthesis. In the book of Chen and Peace (2013), Chapter 8 was devoted to this fundamental question with a simulation study and real data to demonstrate this equivalence. The conclusions from all these research demonstrated that IPD-MA has no gain in statistical efficiency, at least asymptotically, over SS-MA for fixed-effects meta-analysis models.

However, fixed-effects MA models rely on a critical and very strong assumption that the effect-sizes from all studies share a common value across all studies. This is not true in many MAs since it is not practical to assume that all studies will be homogeneous since they are from very diversified sources. It is thus more appropriate to use random-effects MA models which assume that study-specific effects are realizations of a random variable. Although random-effects MA models are widely used in the literature as well as in practice, the fundamental question on the relative efficiency from IPD-MA to SS-MA remains unanswered for this important class of models. In this chapter, a simulation study is designed to demonstrate this relative equivalence under different scenarios from small to large study-specific sample sizes as well as small to large number of studies in meta-analysis.

This chapter is then organized as follows. In Sect. 2, we give an extensive overview on meta-analysis using summary statistics. Specifically, we review the

classical SS-MA on (1) fixed-effects and random-effects MAs, (2) quantification of heterogeneity with Q-statistic, τ^2 index, H index and I^2 index, and (3) meta-regression to explore the heterogeneity with study-level moderators. With the knowledge on SS-MA, Sect. 3 further demonstrate the relative efficiency between IPD-MA and SS-MA with a simulation study under continuous and binary data. Section 4 will illustrate the meta-analysis with a real data set from 13 clinical trials conducted to assess the impact of a Bacillus Calmette-Guerin (BCG) vaccine in the prevention of tuberculosis (TB). Detailed analysis using open source R software is illustrated in this section so that the interested readers can follow the meta-analysis form their own real data analysis. We conclude this chapter with a discussion in Sect. 5.

2 Overview of Meta-Analysis with Summary Statistics

2.1 Summary Statistics and the Sources of Variations

In a typical meta-analysis with summary statistics (SS-MA), K independent studies are obtained to estimate a parameter of interest, such as the effect-size of efficacy between new treatment and control, β_k $(k = 1, 2, ..., K)$. This can be referred to a broad range of study designs such as single-arm or multiple-arm studies, randomized controlled studies, and observational studies. For easy illustration, we focus on SS-MA with two-arm studies, where β_k is some form of the effect-size between the two groups. The most popular choice for β_k is the mean difference or the standardized mean difference for a continuous outcome, or odds-ratio, risk ratio, and risk difference for dichotomous outcome. In most cases, an estimate $\widehat{\beta}_k$ of the true β_k and its associated standard error could be directly extracted from each study. The ultimate goal of meta-analysis is to produce an optimal estimate of the overall population effect-size by pooling the estimates $\widehat{\beta}_k$ $(k = 1, 2, ..., K)$ from individual studies using appropriate statistical models.

Two sources of errors or variations existed in these summary statistics of $\widehat{\beta}_k$ from different studies with one as the within-study variation and another as the between-study variation. The within-study variation is caused by sampling error, which is random or non-systematic from each study. In contrast, the between-study variation is resulted from the systematic differences among studies. If the between-study variation can be verified to be zero, the effect estimates $\widehat{\beta}_k$ are considered homogeneous. Otherwise, they are heterogeneous. In SS-MA, the assumption of homogeneity states that β_k $(k = 1, 2, ..., K)$ are the same in all studies, that is

$$\beta_1 = \beta_2 = \cdots = \beta_K = \beta. \tag{1}$$

Based on whether these studies are homogeneous, two commonly used meta-analysis models can be seen in SS-MA with one as fixed-effects MA models and another as random-effects MA models.

2.2 Fixed-Effects Meta-Analysis

The fixed-effects meta-analysis assumes homogeneous where the underlying population effect-sizes β_k are constant across all studies as shown in Eq. (1). A typical fixed-effects model is described as

$$\widehat{\beta}_k = \beta + \epsilon_k;\ k = 1, 2, ..., K, \tag{2}$$

where $\widehat{\beta}_k$ represents the effect-size for study k and β is the global overall population effect-size. The ϵ_k are the sampling error with mean 0 and KNOWN variance $\widehat{\sigma}^2_{\beta_k}$ which can be extracted or calculated from the individual studies. In general, the ϵ_k is assumed to follow a normal distribution $N(0, \widehat{\sigma}^2_{\beta_k})$. A pooled estimate of β in fixed-effects MA models is given by the weighted least square estimation

$$\widehat{\beta}_F = \frac{\Sigma_{k=1}^K w_k \widehat{\beta}_k}{\Sigma_{k=1}^K w_k}, \tag{3}$$

and the variance of $\widehat{\beta}_F$ can be expressed as

$$Var\left(\widehat{\beta}_F\right) = 1/\Sigma_{k=1}^K w_k, \tag{4}$$

where a popular choice of weight is $w_k = 1/\widehat{\sigma}^2_{\beta_k}$ and variance $\widehat{\sigma}^2_{\beta_k}$ is estimated from study k. Hence, the 95% confidence interval of β_F is given by

$$\widehat{\beta}_F - t_{0.025,(K-1)}\sqrt{Var\left(\widehat{\beta}_F\right)} \le \beta \le \widehat{\beta}_F + t_{0.025,(K-1)}\sqrt{Var\left(\widehat{\beta}_F\right)}, \tag{5}$$

where $t_{0.025,(K-1)}$ denotes the 2.5%-percentile of a $t-$distribution with $(K-1)$ degrees of freedom. Similarly, we may formulate a statistical t-test as:

$$t = \frac{\widehat{\beta}_F - \beta}{\sqrt{Var\left(\widehat{\beta}_F\right)}} \tag{6}$$

to be used to test $H_0 : \beta = 0$.

2.3 Random-Effects Meta-Analysis

When meta-analyzing effect-sizes from different studies, the assumption in the fixed-effects model that the K true effect-sizes are the same for all studies may be impractical. When we attempt to synthesize a group of studies with a meta-analysis, we expect that these studies have enough in common to combine the information for statistical inference, but it would be impractical to require that these studies have identical true effect-sizes. It is impossible for independent studies to be identical in every respect. Therefore heterogeneity should be very likely to exist in many MAs. The model that takes heterogeneity into account is the following random-effects meta-analysis models:

$$\widehat{\beta_k} = \beta + b_k + \epsilon_k, k = 1, 2, ..., K, \tag{7}$$

where for study k, $\widehat{\beta_k}$ represents the observed effect-size and β the global population effect-size. b_k is now the random-effects with mean 0 and variance τ^2 representing the between-study heterogeneity, and ϵ_k is the sampling error with mean 0 and variance $\widehat{\sigma}^2_{\beta_k}$. It is assumed that b_k and ϵ_k are independent and follow normal distributions $N(0, \tau^2)$ and $N(0, \widehat{\sigma}^2_{\beta_k})$, respectively. Let $\beta_k = \beta + b_k$, $k = 1, 2, ..., K$. Then the random-effects model (7) can be simplified as

$$\widehat{\beta_k} = \beta_k + \epsilon_k, \tag{8}$$

where β_k represents the true effect-size for study k. All β_k $(k = 1, 2, ..., K)$ are random samples from the same normal population

$$\beta_k \sim N(\beta, \tau^2) \tag{9}$$

rather than being a constant in the fixed-effects MA in Eq. (2.2). Further, the marginal variance of $\widehat{\beta_k}$ is given by

$$Var\left(\widehat{\beta_k}\right) = \tau^2 + \widehat{\sigma}^2_{\beta_k}, \tag{10}$$

which is composed of two sources of variation, i.e., the between-study variance τ^2 and within-study variance $\widehat{\sigma}^2_{\beta_k}$. If the between-study variance $\tau^2 = 0$, the random-effects MA models (7) would reduce to the fixed-effects MA models (2).

Similar to the fixed-effects MA models, the within-study variance $\widehat{\sigma}^2_{\beta_k}$ can be obtained or calculated from each study k $(k = 1, 2, ..., K)$. However, information of between-study variance τ^2 is often not available and methods commonly used for assessing between-study heterogeneity include the DerSimonian-Laird's method of moments (MM) in DerSimonian and Laird (1986), the maximum likelihood estimation (MLE) method in Hardy and Thompson (1998), and the restricted maximum likelihood (REML) method in Raudenbush and Bryk (1985). As the most commonly used estimator, MM is a distribution-free and non-iterative approach

whereas both MLE and REML are parametric methods and need iteration for estimating τ^2 and the related meta-regression parameter β described in Sect. 2.5 below.

For MM, it utilizes the Q-statistic that is used for testing the assumption of homogeneity:

$$Q = \Sigma_{k=1}^{K} w_k \left(\widehat{\beta}_k - \widehat{\beta}_F\right)^2, \tag{11}$$

where $w_k = 1/\widehat{\sigma}^2_{\beta_k}$ is the weight from the kth study, $\widehat{\beta}_k$ is the kth study effect-size, and $\widehat{\beta}_F$ is the global overall effect estimated from Eq. (3). It can be seen that Q is calculated as: (1) compute the deviations of each effect-size from the meta-estimate and square them (i.e., $(\widehat{\beta}_k - \widehat{\beta}_F)^2$), (2) weight these values by the inverse-variance for each study, and (3) then sum these values across all K studies to produce a weighted sum of squares (WSS) to obtain the heterogeneity measure Q.

From Eq. (11), Chen and Peace (2013) has shown that the expected value of Q is

$$E(Q) = (K-1) + U \times \tau^2, \tag{12}$$

where $U = \sum_{k=1}^{K} w_k - \frac{\sum_{k=1}^{K} w_k^2}{\sum_{k=1}^{K} w_k}$. Under the assumption of no heterogeneity (all studies have the same effect-size), then τ^2 would be zero and $E(Q) = df = K-1$. Based on this heterogeneity measure Q, the test of heterogeneity is conducted to address the null hypothesis that the effect-sizes β_k from all studies share a common effect-size β (i.e., the assumption of homogeneity) and then test this hypothesis where the test statistic is constructed using Q as a central χ^2 distribution with degrees of freedom of $df = K-1$. It should be cautioned that this χ^2-test using the Q-statistic has poor statistical power to detect true heterogeneity for a meta-analysis with a small number of studies K, but excessive power to detect negligible variability with a large number of studies as discussed in Harwell (1997) and Hardy and Thompson (1998). Thus, a nonsignificant test using Q-statistic from a small number of studies can lead to an erroneous selection of a fixed-effects model when there is possible true heterogeneity among the studies, and vice versa. The inability to conclude statistically significant heterogeneity in a meta-analysis of a small number of studies at the 0.05 level of significance is similar to failing to detect statistically significant treatment-by-center interaction in MRCT. In these settings, many analysts will conduct the test of homogeneity at the 0.10 level, as a means of increasing power of the test.

From Eq. (12), the MM estimate of τ^2 can be shown (Chen and Peace, 2013) as follows

$$\widehat{\tau}^2 = \max\left(0, \frac{Q-(K-1)}{U}\right). \tag{13}$$

The truncation at zero in (13) is to ensure that the variance estimate is non-negative.

Therefore, the estimate of the global population effect-size in random-effects MA is given by

$$\widehat{\beta}_R = \frac{\Sigma_{k=1}^{K} w_k^* \widehat{\beta}_k}{\Sigma_{k=1}^{K} w_k^*}, \tag{14}$$

where $w_k^* = 1/\left(\widehat{\sigma}_{\beta_k}^2 + \widehat{\tau}^2\right)$. The variance of $\widehat{\beta}_R$ is simply

$$Var\left(\widehat{\beta}_R\right) = 1/\Sigma_{k=1}^{K} w_k^*$$

and the 95% confidence interval can be constructed by $\widehat{\beta}_R - t_{0.025,(K-1)}\sqrt{Var\left(\widehat{\beta}_R\right)}$ $\leq \beta \leq \widehat{\beta}_R + t_{0.025,(K-1)}\sqrt{Var\left(\widehat{\beta}_R\right)}$.

The statistical test can be similarly formulated by:

$$t = \frac{\widehat{\beta}_R - \beta}{\sqrt{Var\left(\widehat{\beta}_R\right)}} \tag{15}$$

to be used to test $H_0 : \beta = 0$ in the random-effects MA framework.

2.4 Quantify Heterogeneity in Meta-Analysis

The Q-statistic can be used to test the existence of heterogeneity, but it does not report the extent of such heterogeneity. Based on this Q-statistic, several indices are proposed to quantify heterogeneity, such as: τ^2, H, and I^2.

2.4.1 The τ^2 Index

The τ^2 index is defined as the variance of the true effect-size as seen in the random-effects MA model (13). Since it is impossible to observe the true effect-size, we cannot calculate this variance directly, but we can estimate it from the observed data using Eq. (11) as follows:

$$\widehat{\tau}^2 = \frac{Q - (K-1)}{U} \tag{16}$$

which is the well-known DerSimonian-Laird method of moments for τ^2 in Eq. (13). Even though the true variance τ^2 can never be less than zero, the estimated variance $\widehat{\tau}^2$ can sometimes be from the sampling error leading to $Q < K - 1$. When this occurs, the estimated $\widehat{\tau}^2$ is set to zero.

As used in the random-effects MA model, the τ^2 index is also an estimate for the between-studies variance in the meta-analysis.

2.4.2 The H Index

Another index or measure of heterogeneity is the H, proposed in Higgins and Thompson (2002), and defined as follows:

$$H = \sqrt{\frac{Q}{K-1}}. \tag{17}$$

This index is based on the fact that $E[Q] = K-1$ when there is no heterogeneity. In this case, H should be 1.

The confidence interval for the H index is derived in Higgins and Thompson (2002) based on the assumption that the natural logarithm of $\ln(H)$ follows a standard normal distribution. Accordingly:

$$LL_H = \exp\left\{\ln(H) - |z_{\alpha/2}| \times SE\left[\ln(H)\right]\right\} \tag{18}$$

$$UL_H = \exp\left\{\ln(H) + |z_{\alpha/2}| \times SE\left[\ln(H)\right]\right\}, \tag{19}$$

where LL and UL denote the lower- and upper-limits of the CI, $z_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution, and $SE[\ln(H)]$ is the standard error of $\ln(H)$ and is estimated by

$$SE\left[\ln(H)\right] = \begin{cases} \frac{1}{2}\frac{\ln(Q)-\ln(K-1)}{\sqrt{2Q-\sqrt{2K-3}}} & \text{if } Q > K \\ \sqrt{\frac{1}{2(K-2)}\left(1-\frac{1}{3(K-2)^2}\right)} & \text{if } Q \le K \end{cases}. \tag{20}$$

Since $E(Q) \approx K-1$ as seen Eq. (12), the H index should be greater than 1 to measure the relative magnitude of heterogeneity among all the studies. If the lower limit of this interval is greater than 1, the H is statistically significant.

2.4.3 The I^2 Index

To measure the proportion of observed heterogeneity from the real heterogeneity, Higgins and Thompson (2002) and Higgins et al. (2003) proposed the I^2 index as follows:

$$I^2 = \left(\frac{Q-(K-1)}{Q}\right) \times 100\%, \tag{21}$$

which again represents the ratio of excess dispersion to total dispersion and is similar to the well-known R^2 in classical regression which represents the proportion of the total variance that can be explained by the regression variables.

As suggested from Higgins et al. (2003) a value of the I^2 index around 25%, 50%, and 75% could be considered as *low-, moderate-*, and *high*-heterogeneity, respectively. As noted in their paper, about half of the meta-analyses of clinical trials in the *Cochrane Database of Systematic Reviews* reported an I^2 index of zero and the rest reported evenly distributed I^2 indices between 0% and 100%.

Mathematically, the I^2 index can be represented using the H index as follows:

$$I^2 = \frac{H^2 - 1}{H^2} \times 100\%. \tag{22}$$

This expression allows us to use the results from the H index to give a confidence interval for the I^2 index using the expressions in Eqs. (18) and (19) as follows:

$$LL_{I^2} = \left[\frac{(LL_H)^2 - 1}{(LL_H)^2}\right] \times 100\%$$

$$UL_{I^2} = \left[\frac{(UL_H)^2 - 1}{(UL_H)^2}\right] \times 100\%.$$

Since I^2 represents the percentage, any of these limits which is computed as negative is set to zero. In the case that the lower limit of I^2 is greater than zero, then the I^2 is regarded as statistically significant.

2.5 Meta-Regression

Random-effects meta-analysis can take into account of between-study heterogeneity, but it is not a tool to explain and explore how these heterogeneities are originated with other study-level variables. When there are study-level variables (which is commonly called "moderators") available, the meta-regression is then used to investigate the association between these moderators and the reported effect-sizes. With study-level moderators associated with the reported effect-sizes as the dependent variable and their variance as weights, meta-regression is in fact the typical weighted regression. From this point of view, meta-regression is merely typical multiple regression applied for study-level data and therefore the theory of regression can be directly applied for meta-regression.

We introduce two types of meta-regressions, which are built on the fixed-effects and random-effects MA models, respectively. Suppose that there are p moderators $X_1, X_2, ..., X_p$ and K independent studies. The fixed-effects meta-regression model is given by

$$\widehat{\beta_k} = \beta_0 + \beta_1 x_{k1} + ... + \beta_p x_{kp} + \epsilon_k, \tag{23}$$

where $x_{k1}, ..., x_{kp}$ $(k = 1, 2, ..., K)$ denote the observed values of the p moderators $X_1, X_2, ..., X_p$ for study k and $\beta_0, \beta_1, ..., \beta_p$ are regression coefficients. The effect-size $\widehat{\beta_k}$ and sampling error ϵ_k have the same definitions as in fixed-effects MA models in (2). The model assumes that the variation in effect-sizes can be completely explained by these moderators. In other words, the variation is predictable and hence the fixed-effects meta-regression.

The random-effects meta-regression can be obtained by adding random-effects b_k to the fixed-effects model in (23):

$$\widehat{\beta_k} = \beta_0 + \beta_1 x_{k1} + ... + \beta_p x_{kp} + b_k + \epsilon_k, k = 1, 2, ..., K,$$

where b_k is assumed independent with a mean 0 and variance τ^2. Unlike the fixed-effects meta-regression, where all variability in effect-sizes can be explained by the moderators $X_1, X_2, ..., X_p$, the random-effects meta-regression assumes that the model can explain only part of the variation and random-effects b_k account for the remainder.

It should be noted that the meta-regression technique is most appropriate when the number of studies in a MA is large. Furthermore, since the moderators and outcome in meta-regression are all study-level summary statistics (e.g., patient mean age, proportion of female patients), the relation between these moderators and outcome may not directly reflect the relation between subject scores and subjects' outcomes, causing aggregation bias. Therefore careful consideration and interpretation of the results are always recommended when performing meta-regression.

3 Simulation Study on Relative Efficiency Between SS-MA and IPD-MA

We now present simulation study to demonstrate the relative efficiency between meta-analysis using SS and IPD to address the fundamental question on the relative efficiency from IPD-MA to SS-MA. It has been demonstrated that for fixed-effects MA, the IPD-MA has no gain in statistical efficiency, at least asymptotically, over SS-MA as outlined in Sect. 1. However, fixed-effects MA models heavily rely on the critical and very strong assumption that the effect-sizes from all studies are homogeneous with a common value across all studies which is not true in many meta-analyses. It is thus more appropriate to use random-effects MA models to incorporate the between-study heterogeneity.

In this sense, the random-effects MA models should be more commonly used in the literature as well as in practice and the fundamental question on the relative efficiency from IPD-MA to SS-MA would be more central and important in

meta-analysis. However, as Lin and Zeng (2010) stated, "it is technically more challenging to deal with random-effects models than fixed-effects models" and until now this is still an unanswered question in the meta-analysis due to the high-dimension of integration. In this section, we design a simulation study to demonstrate this relative equivalence under different scenarios in random-effects meta-analysis for both continuous and binary outcomes.

3.1 Continuous Data

3.1.1 Simulation Settings

In this simulation study, we consider a general data setting for random-effects MA models with K independent studies. Assume for each study, there are n_k individuals from the kth study and let $(\boldsymbol{Y}_{ki}, \boldsymbol{X}_{ki}), k = 1, \cdots, K; i = 1, \cdots, n_k$ denote the original individual patient-level data (IPD) for the n_k individuals in the k-th study $(k = 1, \ldots, K)$ where $\boldsymbol{Y}_{ki}$ are the response variable which can be continuous or discrete, univariate or multivariate, and $\boldsymbol{X}_{ki}$ are the corresponding vector of explanatory variables including the treatment indicators.

For continuous IPD, the meta-analysis from typical clinical trials to compare "Treatment" to "Control" groups can be easily formulated as follows:

$$Y_{ki} = \alpha_k + \boldsymbol{\beta}_k X_{ki} + \epsilon_{ki}, \quad \epsilon_{ki} \sim N(0, \sigma_k^2), \tag{24}$$

where $\boldsymbol{\beta}_k$ represents the random-effects of treatment effect from the kth study and X_{ki} as treatment indicators. The random-effects $\boldsymbol{\beta}_k$ are of primary interest and assumed to be $\boldsymbol{\beta}_k \mid \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \tau_\beta^2)$ where β is the global overall treatment effect. The nuisance parameters of α_k and σ_k^2 are specific to the k-th study.

Specifically, the simulation is conducted in following three steps:

1. **Data Generation Step**:

   Data $(\boldsymbol{Y}_{ki}, \boldsymbol{X}_{ki}), k = 1, \cdots, K; i = 1, \cdots, n_k$ are generated based on Eq. (24) where we simulated ϵ_{ki} from $N(0, \sigma_k^2)$ with all $\sigma_k = 2$. The random-effects β_k are simulated from $N(\beta, \tau_\beta^2)$ with $\beta = 1$ and $\tau_\beta = 1$. X_{ki} are the treatment assignment with "Treatment" and "Control" to mimic a standard two-treatment clinical trial with probability of 50% to assign an individual to each treatment in each study.

   The nuisance parameters of α_k are also simulated from the normal distribution of $\alpha_k \sim N(\alpha, \tau_\alpha^2)$ with $\alpha = 1$ and $\tau_\alpha = 1$. More realistically, we assume that there is correlation of $\tau_{\alpha\beta} = 50\%$ between the random-effects of α_k and β_k. The sample size n_k from all K studies are simulated from a Poisson distribution with the average sample size of n. With all α_k, β_k, X_{ki}, and ϵ_{ki} generated, the y_{ki} can be calculated from the Eq. (24).

2. **IPD-MA Step**:
   For the data generated from the Step 1, the parameter estimation from meta-analysis using IPD (i.e., IPD-MA) can be done using the linear mixed-effects model in Eq. (24) where the random-effects $\alpha_k \sim N(\alpha, \tau_\alpha^2)$ and $\beta_k \sim N(\beta, \tau_\beta^2)$ with α and β as the global overall effects. In these two random-effects, τ_α^2 and τ_β^2 are the between-study heterogeneity and these two random-effects can be correlated too with correlation of $\tau_{\alpha\beta}$.

   Even the primary interest is to estimate the global effect of β, the IPD-MA can simultaneously estimate other nuisance parameters, such as, the between-study heterogeneity from the intercept parameter α and a global within-study variance σ^2 from ϵ_{ki}. The IPD-MA is in fact a 2-level hierarchical linear model with the first-level to include the IPD data from all patients and the second-level to include the K studies and therefore, the IPD-MA is always regarded as the gold standard in statistical modelling. The parameter estimation and the statistical inference are implemented in R with the function *lmer* in R package *lme4* which is created by Bates Douglas and his research group and maintained by Ben Bolker (http://www.bbolker+lme4@gmail.com) for "Linear Mixed-Effects Models using 'Eigen' and S4."
3. **SS-MA Step**:
   For the data generated from the Step 1, the SS-MA can be traditionally carried out in two ways as follows

   - **SS-MA1**:
     The SS-MA1 is to meta-analyze the estimates of the treatment effects $\widehat{\beta}_k$ and $\tau_{\beta_k}^2$ from each study. In this case, the standard linear regression can be used to estimate the treatment effect of β_k from each study $k, k = 1, \cdots, K$ to produce the summary statistics (SS) of treatment effects $\widehat{\beta}_k$ and its variance $\widehat{\tau}_{\beta_k}$. These SS from each study are then used for the random-effects SS-MA as described in Sect. 2.3 to estimate the global treatment effect of β using the Dersimonian-Laird estimator as follows:

$$\widehat{\beta}_R = \frac{\Sigma_{k=1}^K w_k^* \widehat{\beta}_k}{\Sigma_{k=1}^K w_k^*}, \tag{25}$$

     where $w_k^* = 1/\left(\widehat{\sigma}_{\beta_k}^2 + \widehat{\tau}^2\right)$. The between-study heterogeneity $\widehat{\tau}^2$ is estimated using the Dersimonian-Laird estimator in Eq. (16).
   - **SS-MA2**:
     SS-MA2 is to meta-analyze the data summary aggregated from the first-level individual-patient data into study-level commonly seen in the real applications. In this case, the IPD from each study are aggregated to produce the study-level data in the format of means and standard deviations along with the number of observations from each study by the treatment and control groups. These summary statistics can be denoted by $(n_{T,k}, \bar{X}_{T,k}, S_{T,k}^2)$ for treatment group and $(n_{P,k}, \bar{X}_{P,k}, S_{P,k}^2)$ for placebo group for study

Table 1 Parameter estimation of IPD-MA, SS-MA1, and SS-MA2 with continuous data

		IPD-MA				SS-MA1				SS-MA2			
K	n	Est	SE	ESE	CP	Est	SE	ESE	CP	Est	SE	ESE	CP
3	20	1.013	0.782	0.810	0.912	1.013	0.787	0.765	0.886	1.013	0.788	0.763	0.885
10	20	0.997	0.421	0.431	0.941	0.997	0.426	0.424	0.926	0.998	0.428	0.425	0.924
50	20	0.999	0.192	0.191	0.945	1.000	0.194	0.193	0.944	1.000	0.195	0.194	0.944
100	20	1.001	0.135	0.135	0.947	1.000	0.136	0.137	0.946	1.000	0.137	0.138	0.949
3	50	0.994	0.663	0.636	0.873	0.994	0.663	0.609	0.852	0.994	0.663	0.609	0.852
10	50	1.000	0.361	0.357	0.927	1.000	0.361	0.355	0.922	1.000	0.361	0.355	0.922
50	50	0.998	0.164	0.162	0.942	0.998	0.164	0.163	0.943	0.998	0.164	0.163	0.943
100	50	1.001	0.116	0.115	0.947	1.001	0.116	0.115	0.947	1.001	0.116	0.115	0.947
3	100	1.001	0.625	0.578	0.847	1.001	0.625	0.563	0.832	1.001	0.625	0.563	0.831
10	100	1.001	0.342	0.331	0.913	1.001	0.342	0.331	0.911	1.001	0.342	0.331	0.912
50	100	0.999	0.152	0.152	0.946	0.999	0.152	0.152	0.946	0.999	0.152	0.152	0.946
100	100	0.999	0.109	0.108	0.945	0.999	0.109	0.108	0.945	0.999	0.109	0.108	0.945

$k = 1, \cdots, K$, respectively. The treatment effects for each study k can be estimated by their sample mean differences $\widehat{\beta}_k = \bar{X}_{T,k} - \bar{X}_{P,k}$ with pooled variance estimates of $\widehat{\tau}^2_{\beta_k} = \frac{S^2_{T,k}}{n_{T,k}} + \frac{S^2_{P,k}}{n_{P,k}}$. Then the random-effects MA models can be done using the same formula in Eq. (25).

Both SS-MA1 and SS-MA2 can be carried out using function *rma* in R package *metafor* which is created by Wolfgang Viechtbauer (http://www.wvb@metafor-project.org) as a comprehensive "Meta-Analysis Package for R".

The above 3 steps are simulated 10,000 times for each specifications of $K = 3, 10, 50, 100$ from small to large number of studies and $n = 20, 50, 100$ from small to large sample size in each study. A sampling distribution of $\widehat{\beta}$ can then be generated and the performance can be seen from the mean, standard errors, and coverage probability as well as the relative efficiency.

3.1.2 Results

Table 1 summarizes the β estimation from these 10,000 simulation runs for the three meta-analysis methods of IPD-MA, SS-MA1, and SS-MA2. In this table, the columns "Est" and "SE" are the average of parameter estimates $\widehat{\beta}$ and its standard errors from these 10,000 simulations which should be the consistent estimates for the global overall effect of β and its standard deviation of τ^2_β. The column "ESE" is the so-called empirical SE which is calculated from the average of the standard errors from $\widehat{\beta}$ and the column "CP" is the coverage probability on whether the empirical 95% confidence interval covers the true β which would have to be close to 95%.

Table 2 Relative efficiency between SS-MA1, SS-MA2 and IPD-MA with continuous data

		SS-MA1 to IPD-MA			SS-MA2 to IPD-MA		
K	n	Lower	Median	Upper	Lower	Median	Upper
3	20	0.528	0.889	1.496	0.510	0.877	1.508
10	20	0.676	0.962	1.322	0.661	0.966	1.363
50	20	0.893	1.024	1.199	0.887	1.029	1.228
100	20	0.933	1.031	1.154	0.930	1.037	1.179
3	50	0.568	0.922	1.225	0.566	0.920	1.232
10	50	0.798	0.993	1.176	0.793	0.993	1.183
50	50	0.931	1.006	1.098	0.929	1.006	1.101
100	50	0.953	1.008	1.072	0.952	1.009	1.076
3	100	0.602	0.956	1.134	0.602	0.955	1.136
10	100	0.889	0.999	1.119	0.887	0.999	1.121
50	100	0.948	1.002	1.062	0.948	1.002	1.064
100	100	0.964	1.002	1.043	0.963	1.002	1.044

It can be seen from this table that the $\widehat{\beta}$s from all three methods are unbiased estimates to the true $\beta = 1$ as seen from the columns of "Est" for all combinations of K and n. They are virtually identical from the three methods. As n or K increases, the accuracy (i.e., unbiasedness) of the estimate increases. The "SE" and "ESE" are also very close in all the combinations, especially as n or K increases which leads to a closer 95% CP.

The relative efficiency as defined by the ratios of the empirical variances between SS-MA1, SS-MA2 and IPD-MA can be seen from Table 2 for the 10,000 simulations. Two calculations are made in this table with one for the relative efficiency between SS-MA1 and IPD-MA which is labeled as "SS-MA1 to IPD-MA" and another for the relative efficiency between SS-MA2 and IPD-MA which is labeled as "SS-MA2 to IPD-MA," Within each calculation, we report the 2.5%, 50% and 97.5% quantiles from these 10,000 simulations which are labeled as "Lower," "Median," and "Upper" in the table.

It can be seen from this table that for all cases of n and K, the 95% CIs cover the median relative efficiency where all "Lower" bounds are smaller than 1 and "Upper" bounds larger than 1 which means that the relative efficiency between SS-MA and IPD-MA are statistically equivalent. In further examination, we can see an interesting result that when K is small at $K = 3$ and $K = 10$, the relative efficiencies between MA-SS and MA-IPD are less than 1 which indicate that both MA-SSs are in fact more efficient than the IPD-MA averagely. However, as K and n increase, the medians are closer to 1 indicating both SS-MAs are equivalent to IPD-MA.

It should be noted that the same sizes of $n = 20, 30, 100$ are in fact all really small sample sizes in meta-analysis. We simulated some scenarios with larger sample sizes of $n = 300$ and $n = 500$ and found that the relative efficiencies between these two SS-MAs and IPD-MA are virtually identical.

3.2 Categorical Data

3.2.1 Simulation Settings

Categorical data can be generated similarly and in this section we focus on binary data for studies with binary outcomes. In this case, $\boldsymbol{Y}_{ki}$ are the binary response variable and the model is as follows:

$$P(Y_{ki} = 1|X_{ki}) = \frac{\exp(\alpha_k + \beta_k X_{ki})}{1 + \exp(\alpha_k + \beta_k X_{ki})} \quad (k = 1, \cdots, K; i = 1, \cdots, n_k), \tag{26}$$

where $\boldsymbol{\beta}_k$ represents the random-effects of treatment effect from the kth studies and X_{ki} as treatment indicators. The random-effects $\boldsymbol{\beta}_k$ are of primary interest and assumed to be $\boldsymbol{\beta}_k \mid \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \tau_\beta^2)$. The nuisance parameters of α_k are specific to the k-th study and there is no σ_k^2 in binary data.

The simulation can be conducted in following three steps:

1. **Data Generation Step**:
   Data $(\boldsymbol{Y}_{ki}, \boldsymbol{X}_{ki}), k = 1, \cdots, K; i = 1, \cdots, n_k$ are generated based on Eq. (26). The random-effects treatment effect β_k are simulated from $N(\beta, \tau_\beta^2)$ with $\beta = 1$ and $\tau_\beta = 1$. X_{ki} are the treatment assignment with "Treatment" and "Control" to mimic a standard two-treatment clinical trial with probability of 50% to assign an individual to each treatment in each study.

   The nuisance parameters of α_k are also simulated from the random-effects $\alpha_k \sim N(\alpha, \sigma_\alpha^2)$ with $\alpha = 1$ and $\sigma_\alpha = 1$. The sample size n_k from all K studies are simulated from a Poisson distribution with the average sample size of n. With all α_k, β_k and X_{ki} generated, the y_{ki} can be generated from the Eq. (26) using random sampling from binomial distribution.
2. **IPD-MA Step**:
   For the data generated from the Step 1, the parameter estimation from meta-analysis using IPD (i.e., IPD-MA) can be done using the generalized linear mixed-effects model technique in Eq. (26) where the random-effects $\alpha_k \sim N(\alpha, \tau_\alpha^2)$ and $\beta_k \sim N(\beta, \tau_\beta^2)$ with α and β as the global overall effects. Even the primary interest is to estimate the global effect of β, the IPD-MA can simultaneously estimate other nuisance parameters, such as, the between-study heterogeneity from the intercept parameter α.

   Again this IPD-MA is a 2-level hierarchical generalized linear model with the first-level to include the IPD data from all patients and the second-level to include the K studies and therefore, the IPD-MA is always regarded as the gold standard in statistical modelling. The parameter estimation and the statistical inference are implemented in R with the function *lmer* in R package *lme4*.
3. **SS-MA Step**:
   For the data generated from the Step 1, the SS-MA can be carried out in two ways as follows

Table 3 Nomenclature for 2×2 table of outcome by treatment

	Events	Non-events	Total event
Treatment group	A_k	B_k	$n_{T,k}$
Placebo group	C_k	D_k	$n_{P,k}$

- **SS-MA1**:
  The SS-MA1 is to meta-analyze the estimates of the treatment effects $\widehat{\beta}_k$ and $\tau^2_{\beta_k}$ from each study. In this case, the classical logistic regression can be used to estimate the treatment effects of β_k from each study k, $k = 1, \cdots, K$ to produce the summary statistics (SS) of treatment effects $\widehat{\beta}_k$ and its variance $\widehat{\tau}_{\beta_k}$. These SS from each study are then used for the random-effects SS-MA as described in Sect. 2.3 to estimate the global treatment effect of β using the Dersimonian-Laird estimator as in Eq. (25).
- **SS-MA2**:
  SS-MA2 is to meta-analyze the data summary aggregated from the first-level individual-patient data into study-level commonly seen in the real applications. In this case, the IPD from each study are aggregated to produce the study-level data in the format of 2 by 2 contingency tables to report the number of events and non-events in two groups. Specifically, the data from the kth study can be represented as cells A_k, B_k, C_k, and D_k, as shown in Table 3.

  Corresponding to the classical logistic regression in **SS-MA1**, the odds-ratio (OR) associated with an event is defined as the ratio of the odds of the event in treatment group to the odds of the event in placebo group. It is well-known that the odds of the event for the treatment group in the kth study is $Odds_{T,k} = \frac{p_{T,k}}{1-p_{T,k}} = \frac{A_k}{B_k}$ and the odds of the event for the placebo group in the kth study is $Odds_{P,k} = \frac{p_{P,k}}{1-p_{P,k}} = \frac{C_k}{D_k}$. Then the odds-ratio (OR) of the treatment group to the placebo group for kth study is as follows:

$$OR_k = \frac{Odds_{T,k}}{Odds_{P,k}} = \frac{\frac{A_k}{B_k}}{\frac{C_k}{D_k}} = \frac{A_k D_k}{B_k C_k}. \tag{27}$$

  To approximate the normal distribution in using odds-ratios, we usually convert the odds-ratio to the log-scale and estimate the log odds-ratio and its standard error and use these numbers to perform the meta-analysis. Then we transform the results back into the original metric. With this direction, the log odds-ratio is

$$LogOR_k = ln(OR_k). \tag{28}$$

  The approximate variance can be derived from delta method (Chen and Peace 2013) as follows:

$$var(logOR_k) = \frac{1}{A_k} + \frac{1}{B_k} + \frac{1}{C_k} + \frac{1}{D_k}. \tag{29}$$

Therefore the approximate standard error is:

$$SE_{logOR_k} = \sqrt{V_{logOR_k}}. \tag{30}$$

With these calculations in the log-scale, we then transform them back to original scale for odds-ratios (OR). Then the random-effects MA models can be done using the same formula in Eq. (25).

Similarity, the above 3 steps are simulated 10,000 times for each specifications of $K = 10, 50, 100$ for small to large number of studies and $n = 100, 300, 500$ for small to large sample size in each study. A sampling distribution of $\widehat{\beta}$ can then be generated and the performance can be seen from the mean, standard errors, and coverage probability as well as the relative efficiency.

3.2.2 Results

Table 4 summarizes the β estimation from these 10,000 simulation runs for the three meta-analysis methods of IPD-MA, SS-MA1, and SS-MA2.

It can be seen from this table that the $\widehat{\beta}$s from IPD-MA are unbiased, but biased lower for both SS-MA1 and SS-MA2, especially for SS-MA1. However, the biases diminish as sample size n increases which is consistent with the theory of maximum likelihood estimation. It should be noted that these biases do not change regardless of the number of studies K and remain similar with different K as seen in Table 4. The "SE" and "ESE" are also very close in all the combinations, but the CPs have mixed pattern due to the biases from the β estimation.

The relative efficiency between the SS-MA1, SS-MA2, and IPD-MA can be seen from Table 5 for the 10,000 simulations. It can be seen from this table that for all cases of n (except $n = 20$) and K, the 95% sampling CIs cover the median relative efficiency where all "Lower" bounds are smaller than 1 and "Upper" bounds larger

Table 4 Parameter estimation of IPD-MA, SS-MA1, and SS-MA2 with binary data

		IPD-MA				SS-MA1				SS-MA2			
K	n	Est	SE	ESE	CP	Est	SE	ESE	CP	Est	SE	ESE	CP
10	20	1.013	0.406	0.365	0.907	0.659	0.366	0.473	0.939	0.868	0.339	0.416	0.971
50	20	1.000	0.170	0.170	0.946	0.664	0.157	0.207	0.664	0.863	0.148	0.183	0.929
100	20	0.994	0.121	0.120	0.947	0.659	0.110	0.146	0.312	0.858	0.105	0.129	0.851
10	50	1.003	0.248	0.226	0.904	0.909	0.233	0.268	0.962	0.945	0.235	0.270	0.967
50	50	0.994	0.109	0.107	0.944	0.894	0.101	0.116	0.880	0.924	0.102	0.116	0.926
100	50	0.992	0.077	0.076	0.942	0.891	0.072	0.082	0.759	0.920	0.073	0.081	0.857
10	100	0.998	0.174	0.161	0.903	0.974	0.169	0.189	0.964	0.976	0.169	0.190	0.964
50	100	0.991	0.078	0.076	0.939	0.958	0.075	0.081	0.932	0.960	0.075	0.081	0.934
100	100	0.992	0.054	0.054	0.947	0.958	0.052	0.057	0.901	0.959	0.052	0.057	0.907

Table 5 Relative efficiency between SS-MA1, SS-MA2, and IPD-MA with continuous data

		SS-MA1 to IPD-MA			SS-MA2 to IPD-MA		
K	n	Lower	Median	Upper	Lower	Median	Upper
10	20	0.919	1.682	4.365	0.680	1.298	3.765
50	20	1.102	1.488	2.082	0.830	1.163	1.709
100	20	1.185	1.466	1.855	0.900	1.152	1.509
10	50	0.800	1.343	3.791	0.906	1.335	3.777
50	50	0.867	1.16	1.728	0.923	1.143	1.713
100	50	0.910	1.142	1.500	0.940	1.126	1.482
10	100	0.949	1.274	3.727	0.981	1.275	3.727
50	100	0.932	1.088	1.614	0.954	1.089	1.611
100	100	0.942	1.057	1.395	0.960	1.058	1.395

than 1 which means that the relative efficiency between SS-MA and IPD-MA are statistically equivalent. However, the median relative efficiencies in all combinations are greater than 1 indicating that the IPD-MA is relatively more efficient than the two SS-MAs which is intuitively true since IPD-MA is the gold standard in statistical analysis for the meta-data. The gain of relative efficiencies diminishes as K and n increase as seen in the table that the medians are getting closer to 1 indicating asymptotically the SS-MA1 and SS-MA2 are equivalent to IPD-MA.

4 Real Data Analysis

4.1 Introduction to the Bacillus Calmette-Guerin Vaccine Data

This is a dataset from 13 clinical trials conducted to assess the impact of a Bacillus Calmette-Guerin (BCG) vaccine in the prevention of tuberculosis (TB). The dataset is widely used to illustrate meta-analysis and meta-regression; for example, in the books authored by Everitt and Hothorn (2006) (see Table 12.2), Hartung et al. (2008) (see Table 18.8), Borenstein et al. (2009) (see Table 20.1) and Chen and Peace (2013) (see Table 7.1), as well as in the paper by van Houwelingen et al. (2002) and the R library *metafor* by Viechtbauer (2010). The source dataset was reported in the original publication in Colditz et al. (1994) which included 13 clinical trials of BCG vaccine each investigating the effectiveness of BCG in the treatment of tuberculosis.

It should be noticed that the numbers reported in these references are different even though all of them referenced this dataset as *BCG* with 13 studies from the same publications. The data tables reported from Everitt and Hothorn (2006), Borenstein et al. (2009) and Colditz et al. (1994) are the total number of cases in both BCG and control. However, the dataset reported in the R *metafor* library and van Houwelingen et al. (2002) are the numbers of "negative cases." We will use this

Table 6 Data from studies on efficacy of BCG vaccine for preventing tuberculosis

Trial	*author*	*year*	*tpos*	*tneg*	*cpos*	*cneg*	*ablat*	*alloc*
1	Aronson	1948	4	119	11	128	44	Random
2	Ferguson & Simes	1949	6	300	29	274	55	Random
3	Rosenthal et al.	1960	3	228	11	209	42	Random
4	Hart & Sutherland	1977	62	13, 536	248	12, 619	52	Random
5	Frimodt-Moller et al.	1973	33	5036	47	5761	13	Alternate
6	Stein & Aronson	1953	180	1361	372	1079	44	Alternate
7	Vandiviere et al.	1973	8	2537	10	619	19	Random
8	TPT Madras	1980	505	87, 886	499	87, 892	13	Random
9	Coetzee & Berjak	1968	29	7470	45	7232	27	Random
10	Rosenthal et al.	1961	17	1699	65	1600	42	Systematic
11	Comstock et al.	1974	186	50, 448	141	27, 197	18	Systematic
12	Comstock & Webster	1969	5	2493	3	2338	33	Systematic
13	Comstock et al.	1976	27	16, 886	29	17, 825	33	Systematic

data structure in this chapter which is reproduced here for complete presentation in Table 6.

In this table, *author* denotes the authorship from the 13 studies, *year* is publication year of these 13 studies, *tpos* is the number of TB-positive cases in the BCG vaccinated group, *tneg* is the number of TB-negative cases in the BCG vaccinated group, *cpos* is the number of TB-positive cases in the control group, *cneg* is the number of TB-negative cases in the control group, *ablat* denotes the absolute latitude of the study location (in degrees) and *alloc* denotes the method of treatment allocation with three levels of random, alternate, or systematic assignment.

The purpose of the original meta-analysis was to quantify the efficacy of the BCG vaccine against tuberculosis which was facilitated by a random-effects meta-analysis. It concluded that the BCG vaccine significantly reduced the risk of TB- in the presence of significant heterogeneity. The heterogeneity was explained partially by geographical latitude. In this chapter, we use this dataset to illustrate the meta-analysis procedures in Sect. 2 with implementation in R and also use this data to illustrate the relative efficiency between SS-MA and IPD-MA.

4.2 Fixed-Effects and Random-Effects Meta-Analysis

4.2.1 Effect-Size Calculation

We illustrate the meta-analysis using R package *metafor*. To make use of this package, we first load it into R using the following R code chunk:

```
# load the R library
> library("metafor")
```

```
# load the BCG data
> data("dat.bcg", package = "metafor")
```

For meta-analysis, we need to convert the 2 by 2 contingency table in Table 6 into study-specific effect-size (ES). To be consistent with the IPD-MA, we use the log odds-ratio which can be specified as *measure="OR"* in R function *escalc* which stands for "effect-size calculation" as in following R code chunk:

```
> dat = escalc(measure="OR",ai=tpos,bi=tneg,ci=cpos,
        di=cneg,data=dat.bcg,append = TRUE)
```

The option *append=TRUE* above is to append the calculated ES (i.e., log odds-ratio in Eq. 28 and named as "yi") and the its variance in Eq. 29 (i.e., named "vi") to the original data which can be seen as follows:

trial	author	year	tpos	tneg	cpos	cneg	ablat	yi	vi
1	Aronson	1948	4	119	11	128	44	-0.9387	0.3571
2	Ferguson & Simes	1949	6	300	29	274	55	-1.6662	0.2081
3	Rosenthal et al	1960	3	228	11	209	42	-1.3863	0.4334
4	Hart & Sutherland	1977	62	13536	248	12619	52	-1.4564	0.0203
5	Frimodt-Moller et al	1973	33	5036	47	5761	13	-0.2191	0.0520
6	Stein & Aronson	1953	180	1361	372	1079	44	-0.9581	0.0099
7	Vandiviere et al	1973	8	2537	10	619	19	-1.6338	0.2270
8	TPT Madras	1980	505	87886	499	87892	13	0.0120	0.0040
9	Coetzee & Berjak	1968	29	7470	45	7232	27	-0.4717	0.0570
10	Rosenthal et al	1961	17	1699	65	1600	42	-1.4012	0.0754
11	Comstock et al	1974	186	50448	141	27197	18	-0.3408	0.0125
12	Comstock & Webster	1969	5	2493	3	2338	33	0.4466	0.5342
13	Comstock et al	1976	27	16886	29	17825	33	-0.0173	0.0716

It can be found that 11 of 13 (except the trials 8 and 12) have negative log odds-ratios which indicate that the BCG vaccine reduced the risk of TB for these 11 studies, but not for the trials of 8 and 12. Their statistical significance can be examined by constructing the t-statistic using the ratio of "yi" to the squared-root of "vi" or the 95% confidence intervals as shown in Fig. 1. Examining this figure, we can see that only studies 2, 3, 4, 6, 7, 10, and 11 (i.e., 7 out of 13) are statistically significant.

4.2.2 Fixed-Effects Meta-Analysis

This leads to meta-analysis of all the 13 studies together. Assuming homogenous for all 13 studies, the fixed-effects meta-analysis in Sect. 2.2 can be done using R function "rma" as follows:

```
# Fixed-effects MA
> meta.FE = rma(yi, vi, data = dat, method="FE")
# Print the summary
> meta.FE

Fixed-Effects Model (k = 13)
```

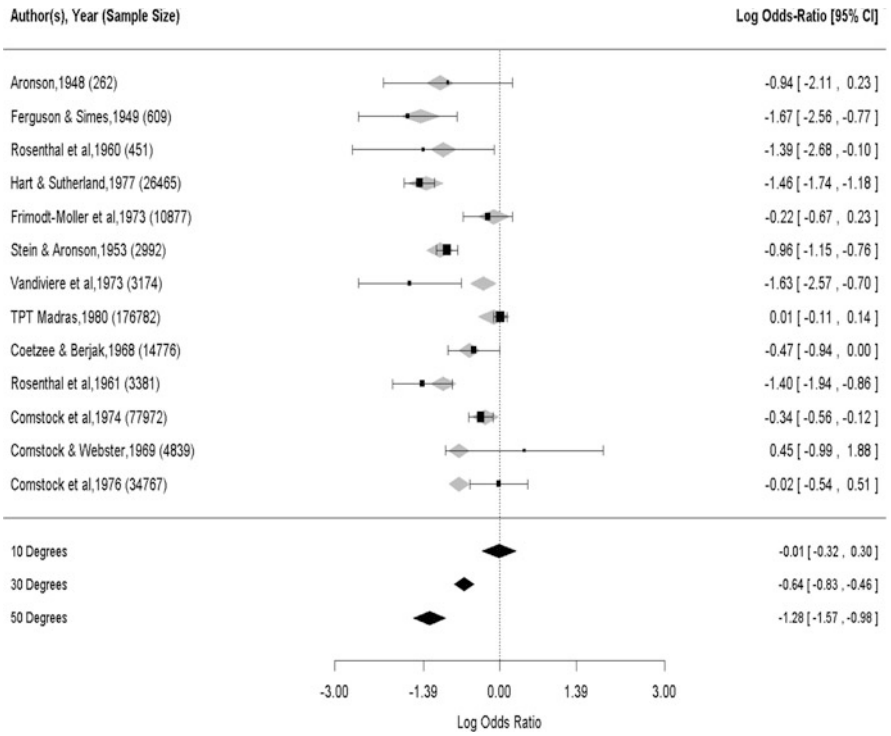


Fig. 1 Forest plot before/after meta-regression

```
Test for Heterogeneity: Q(df = 12) = 163.1649, p-val
< .0001

Model Results:
estimate       se     zval     pval    ci.lb    ci.ub
 -0.4361   0.0423 -10.3190   <.0001  -0.5190  -0.3533
```

This shows that the estimated global log odds-ratio is -0.436 with SE of 0.0423 and p-value <0.0001. This means that when these 13 studies meta-analyzed together, the global overall effect of BCG vaccine is statistically significant in preventing TB. However, the test of heterogeneity (i.e., $Q = 163.1649$ with df $= 12$, p-value <0.0001 from χ^2-test) is highly statistically significant which indicates that the assumption of homogeneous studies in fixed-effects meta-analysis is not appropriate. Then random-effects meta-analysis should be explored which is exactly the **SS-MA2 in Step 3** described in Sect. 3.2.1.

4.2.3 Random-Effects Meta-Analysis

The Dersimonian-Laird random-effects meta-analysis models can be implemented easily as follows:

```
# Call 'rma' with method="DL" to fit the random-
  effects MA
> meta.DL = rma(yi, vi, data = dat, method="DL")
# Print the summary
> meta.DL

Random-Effects Model (k = 13; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.3663
      (SE = 0.2659)
tau (square root of estimated tau^2 value):      0.6053
I^2 (total heterogeneity / total variability):   92.65%
H^2 (total variability / sampling variability):  13.60

Model Results:
estimate       se     zval     pval    ci.lb    ci.ub
 -0.7474   0.1923  -3.8873   0.0001  -1.1242  -0.3706
```

The global overall estimate from the random-effects meta-analysis model is $\widehat{\beta} = -0.747$ with SE $= 0.192$ and p-value $= 0.001$ which again indicated statistical significance if meta-analyzed together. In this random-effects meta-analysis, the estimated between-study heterogeneity $\widehat{\tau}^2 = 0.3663$ with $I^2 = 92.65\%$ and $H^2 = 13.60$ which demonstrated highly and statistically significant heterogeneity. Meta-regression should be resorted to explain this heterogeneity.

4.3 Meta-Regression

To explain the heterogeneity, we make use of the geographical latitude where the clinical trials were conducted. This random-effects meta-regression can be implemented in following R code chunk:

```
# random-effects meta-regression
> metaReg = rma(yi, vi, mods = ~ablat, method="DL",data = dat)
# Print the meta-regression results
> metaReg

Mixed-Effects Model (k = 13; tau^2 estimator: DL)

tau^2 (estimated amount of residual heterogeneity):     0.0480 (SE = 0.0451)
tau (square root of estimated tau^2 value):             0.2191
I^2 (residual heterogeneity / unaccounted variability): 56.17%
H^2 (unaccounted variability / sampling variability):   2.28
R^2 (amount of heterogeneity accounted for):            86.90%

Test for Residual Heterogeneity: QE(df = 11) = 25.0954, p-val = 0.0088
```

```
Test of Moderators (coefficient(s) 2): QM(df = 1) = 26.1628, p-val < .0001

Model Results:
         estimate      se     zval    pval    ci.lb    ci.ub
intrcpt    0.3030  0.2109   1.4370  0.1507  -0.1103   0.7163
ablat     -0.0316  0.0062  -5.1150  <.0001  -0.0437  -0.0195
```

As can be seen from the output, with only *ablat*, the estimated residual heterogeneity $\widehat{\tau}^2$ dropped to 0.048 (SE = 0.0451) from the previous random-effect meta-analysis without *ablat* where the heterogeneity was estimated at 0.3663. This indicates that the moderator *ablat* itself accounts for $(0.3663 - 0.0480)/0.3663 = 87\%$ of the total amount of heterogeneity, and the absolute latitude is significantly related to the effectiveness of the BCG vaccine in preventing TB which can be quantified in the estimated meta-regression equation as follows:

$$log(OR) = 0.3030 - 0.0316 \times ablat. \tag{31}$$

This estimated equation and the entire meta-regression summary can be graphically displayed in Fig. 1. In this figure, the meta-regression in Eq. (31) is plotted in the solid line and its 95% CI bounds are plotted in dashed dots. The horizontal line from OR = 0 indicates no BCG vaccine effects. The 13 studies are plotted in 13 solid dots where the sizes of the dots are proportional to their sample sizes.

It can be seen from this equation and the associated Fig. 2 that the higher the absolute latitude, the more effective is the BCG vaccine. When the *ablat* is less than 20° and close to zero (i.e., study performed closer to equator), the effect-size would be close to zero (as evidenced from the insignificant intercept parameter and Fig. 2 which means that the vaccination has no real effect on prevention TB. As *ablat* increases, say to a latitude of 60°, the log OR as calculated from Eq. (31) is −1.593 which corresponds to a OR of 0.20.

This property is further illustrated in Fig. 1. In this figure, the right side is the point estimate of log odds-ratio and its 95% CIs for each of these 13 studies. The middle part is the graphical representation of these 95% CIs in solid line segments. The shaded diamonds are the 95% CIs after the meta-regression adjustment. The bottom three dark diamonds are the 95% CIs for the random-effects meta-regression at three latitudes of 10, 30, 60°. It can be seen at 10°, the estimated log odds-ratio is −0.01 and the 95% CI is (−0.32, 0.30) indicating no statistically significant benefit of this BCG vaccine.

4.4 Relative Efficiency Between SS-MA and IPD-MA

4.4.1 Individualization Procedure

To use this BCG data for IPD analysis, we first need to "individualize" this meta-data by assigning individual IDs to each patient from each treatment group (i.e., "BCG" and "Control" groups) from each of these 13 trials.

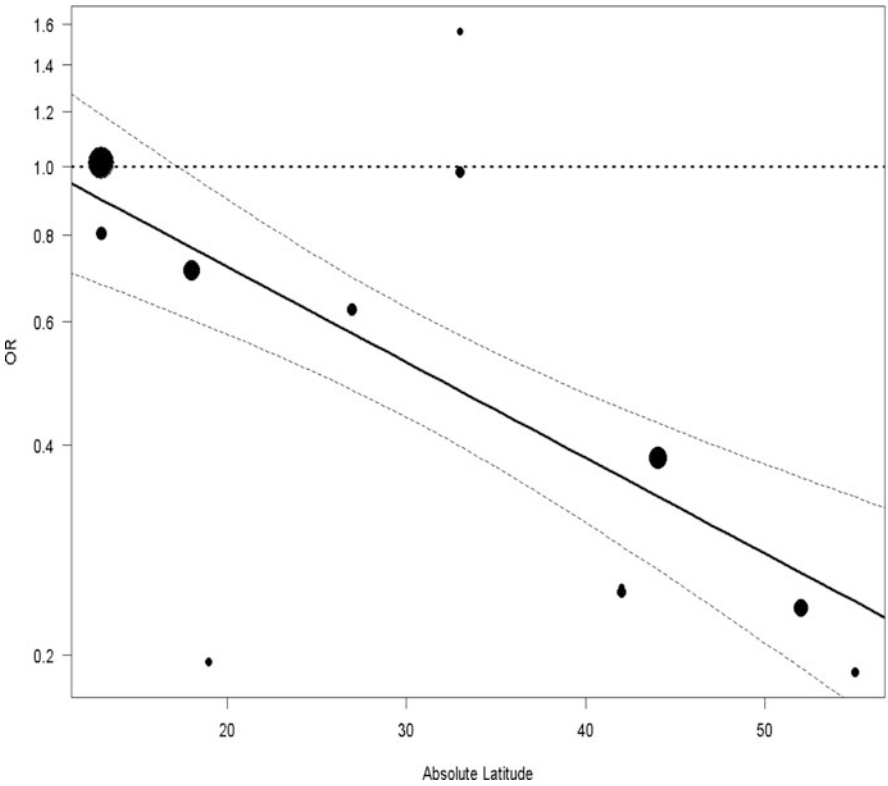


Fig. 2 Random-effects meta-regression for BCG data on OR on absolute latitude

For example, in Table 6, there are 4 TB-positive patients and 119 TB-negatives patients from the BCG group at trial 1. This "individualization" would create a dataframe with variables (1) a patient identification variable (named as "ID") from 1 to 123(=4 + 119); (2) a variable for TB status (named as "TBStatus") assigning the first 4 patients to 1 as TB-positive and the rest 119 to 0 as TB-negative; (3) a group variable (named as "Group") to indicate which treatment group (i.e., BCG) these 123 patients are from and (4) a trial variable (named as "trial") to indicate the trial number (1 for these 123 patients). A random assignment of these 4 deaths to the 119 patients can be done too. Therefore corresponding to these 123 patients, this "individualization" process would create a dataframe with 4 columns with first column (named as "ID") as the patients' ID from 1 to 123, the 2nd column (named as "TBStatus") with the first 4 patients to be 1 and the rest 119 patient to be 0, the third column (named as "Group") as "BCG" for these 123 patients and the fourth column (named as "trial") to denote the clinical trial number and it would be 1 for these 123 patients.

Similarity, we can "individualize" the 139(=11 + 128) patients from the control group to add 139 observations to the previous 123 patients with "ID" from 124 to

262(=123 + 139), "TBStatus" to be 1 for the 11 patient and 0 for the 128 patients, "Group" to be "Control" for these 139 patients and "trial" still to be 1 for these 139 patients. This "individualization" process is very intuitive and can be done for the rest 12 trials which would generate a dataframe with 357,347 observations (i.e., the total patients from the 13 trials as seen from Table 6 and 4 columns.

4.4.2 IPD-MA

With this IPD data, the generalized linear mixed-effects model (i.e., IPD-MA in Step 2 described in Sect. 3.2.1) can be fitted with a binomial distribution to TB status by treatment "Group" assuming random-effects for "trial." With this IPD-MA model fitting, the estimated global overall treatment effect (in log odds-ratio) is -0.7416 with standard error of 0.1201 which gave a significant p-value <0.0001 which means that the BCG vaccine statistically significantly prevented the TB. The estimated heterogeneity standard deviation (between-study standard deviation) is $\widehat{\tau}_{\beta} = 0.5797$.

In addition to the estimation of the global overall treatment effect β, this IPD-MA can also simultaneously estimate the overall intercept in Eq. 26 which is the overall effect for placebo in log odds. The overall log odds for placebo is estimated to be -4.1234 (standard error = 0.1560 and p-value <0.001). Transformed back from the log odds, the odds is then to be 0.0162 which is equivalent to the risk of 1.59% in the probability scale. This means that for people without BCG vaccine, the probability to get TB is about 1.59% which is very small.

4.4.3 SS-MA1

With the IPD data again, we can also perform the SS-MA1 in Step 3 described in Sect. 3.2.1. We can fit the classical logistic regression for each of these 13 studies and keep track of the BCG effects and their estimated variances from this logistic regression for each study which can be shown as follows:

```
          yi     vi
1   -0.9387 0.3571
2   -1.6662 0.2081
3   -1.3863 0.4334
4   -1.4564 0.0203
5   -0.2191 0.0519
6   -0.9581 0.0099
7   -1.6338 0.2270
8    0.0120 0.0040
9   -0.4717 0.0570
10 -1.4012 0.0754
11 -0.3408 0.0125
```

```
12  0.4466 0.5342
13 -0.0173 0.0716
```

It should be noted that these values of estimates are obtained from maximum likelihood estimation in logistic regression. It can be seen that these values of yi and vi from logistic regression are identical to the values from the values in Sect. 4.2 calculated using "escalc" using Eqs. 27 and 28. Therefore the random-effects SS-MA1 and SS-MA2 described in Sect. 3.2.1 would give identical estimates and we then denote both SS-MA1 and SS-MA2 as SS-MA.

4.4.4 Relative Efficiency Between IPD-MA and SS-MA

As summary from the above analyses, the Dersimonian-Laird random-effects meta-analysis (i.e., SS-MA) yielded that the global estimate (in log odds-ratio) of $\widehat{\beta}_{SS-MA} = -0.7474$ with standard error of 0.1923 and p-value <0.0001 whereas the IPD-MA estimated the global overall treatment effect $\widehat{\beta}_{IPD-MA} = -0.7416$ with standard error of 0.1201 and p-value <0.0001.

It can be seen evidently that these two estimates are very close. The standard error for IPD-MA was estimated at 0.1299, but 0.1923 for SS-MA which resulted the relative efficiency between SS-MA and IPD-MA (defined as the ratio of the estimated variance) of 2.19. The estimated between-study heterogeneity is 0.5798 for IPD-MA and 0.6053 for SS-MA, respectively. This conclusion is consistent with the theory of statistics that the IPD-MA incorporated all the information from the data for a simultaneous model fitting which would be relatively more efficient as the golden standard in all statistical modelling.

5 Summary and Discussions

This chapter gave a detailed overview for the classical statistical meta-analysis including the fixed-effects and random-effects meta-analysis models. Linked with quantification of the heterogeneity in meta-analysis, the indices of τ^2, H, and I^2 were introduced along with the meta-regression to take into account of the between-study heterogeneity and variations from all studies in meta-analysis. To facilitate the understanding of meta-analysis techniques and for interested readers to use these models for their own meta-analysis, a real example was illustrated with step-by-step implementation in R on how to do meta-analyses.

With regard to the ongoing debate on whether using original IPD is more beneficial than using summary statistics in meta-analysis, this chapter designed a simulation study to demonstrate their equivalency in random-effects meta-analysis to add to the discussion. It is well-known that it is extreme costly and time-consuming to collect IPD and often impractical and the summary statistics can be obtained much more easily from published literatures. This chapter addressed a

fundamental issue on whether there is efficiency loss from the random-effects meta-analysis using summary statistics to the gold standard of meta-analysis using IPD in simulation study. This simulation study showed numerically that there is little, if at all, relative efficiency loss between these two approaches in meta-analysis.

Further theoretical work to extend the results from Lin and Zeng (2010) to random-effects meta-analysis models as well as more simulation studies should be explored and we are investigating it.

Acknowledgments The author would like to thank the anonymous reviewer for his/her comments which have significantly improved the quality of this chapter. This work is partially supported by the National Research Foundation of South Africa (Grant Number 127727) and the South African National Research Foundation (NRF) and South African Medical Research Council (SAMRC) (South African DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant Number 114613).

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Wiley.

Chen, D. G., & Peace, K. E. (2013). *Applied meta-analysis using R.* Chapman & Hall/CRC.

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature. *Journal of the American Medical Association, 271*, 698–702.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177–188.

Everitt, B., & Hothorn, T. (2006). *A handbook of statistical analyses using R*. Boca Raton, FL: Chapman & Hall/CRC. ISBN 1-584-88539-4.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine, 17*, 841–856.

Hartung, J., Knapp, G., & Sinha, B. K. (2008). *Statistical meta-analysis with applications*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Harwell, M. (1997). An empirical study of hedges homogeneity test. *Psychological Methods, 2*, 219–231.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560.

Lin, D. Y., & Zeng, D. (2010). On the relative efficiency of the using summary statistics versus individual-level data in meta-analysis. *Biometrika, 97*(2), 321–332.

Liu, D., Liu, R., & Xie, M. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. *Journal of the American Statistical Association, 110*(509), 326–340.

Mathew, T., & Nordstrom, K. (1999). On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics, 55*, 1221–1223.

Olkin, I., & Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics, 54*, 317–322.

Pigott, T. D. (2012). *Advances in meta-analysis*. New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical bayes meta-analysis. *Journal of Educational Statistics, 10*(2), 75–98.

Simmonds, M., & Higgins, J. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine, 26*(15), 2982–2999.

Sutton, A. J., & Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*(5), 625–50.

van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Tutorial in biostatistics: Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine, 21*, 589–624.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the *metafor* package. *Journal of Statistical Software, 36*(1), 1–48.

Whitehead, A. (2003). *Meta-analysis of controlled clinical trials*. New York, NY, USA: John Wiley & Sons, Inc.

Meta-Analysis Using R Statistical Software

Nelson Owuor Onyango and Hesborn Otieno Wao

Abstract A systematic review and meta-analysis (SRMA) of results from randomized clinical trials (RCTs) is considered the highest level of evidence in determining comparative effect of health intervention for a given disease or condition. This exercise involves pooling results of relevant published Randomized Controlled Trials (RCTs) to obtain the totality of evidence on specified outcomes of interest. This chapter mainly focuses on meta-analysis of intervention studies. The chapter aims at introducing the following topics in meta-analysis:

1. Statistical methods behind meta-analysis including measures of disease occurrence (e.g., odds ratio [OR], relative risk [RR], and mean difference [MD]); methods for pooling results (e.g., Peto OR, Mantel Haenszel [MH] Statistic and Inverse Variance [IV]); some study designs in clinical trials; measures of heterogeneity; and subgroup analysis
2. Steps involved in meta-analysis of interventions using an open-source software (R statistical software for construction of forest plots and funnel plots, and computation of heterogeneity indices)
3. Meta-regression including illustrative examples using R codes to demonstrate how meta-analysis is conducted for continuous and dichotomous outcomes

Keywords Computer software · Fixed effect · Meta-analysis · Meta-regression · Random effects

N. O. Onyango (✉)
School of Mathematics, University of Nairobi, Nairobi, Kenya
e-mail: onyango@uonbi.ac.ke

H. O. Wao
African Population and Health Research Center, Nairobi, Kenya

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_7

1 Introduction

This chapter presents statistics useful in meta-analysis. The methods in this chapter mainly focus on meta-analysis for intervention studies. Some of the statistical methods addressed include effect measures such as Odds Ratio (OR), Relative Risk (RR), Mean Difference (MD), Standardized Mean Difference (SMD), and Ratio of Means (ROM). Besides, methods for pooling statistical results such as Peto OR, Mantel Haenszel (MH) Statistic, Inverse Variance (IV) Method, and measures of heterogeneity are discussed. Practical examples are presented using mostly datasets available in relevant R packages. The statistical software of choice is R Statistical Software (https://www.R-project.org/). A brief overview of R statistical software is provided before delving into meta-analysis.

1.1 Citation of R Statistical Software

The R software code `citation()` provides the recommended citation for R statistical software.

```
R>citation()
```

To cite R in publications use:

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

A BibTeX entry for LaTeX users is

@Manual,
title = R: A Language and Environment for Statistical Computing,
author = R Core Team,
organization = R Foundation for Statistical Computing,
address = Vienna, Austria,
year = 2019,
url = https://www.R-project.org/,

We have invested a lot of time and effort in creating R, please cite it when using it for data analysis. See also 'citation("pkgname")' for citing R packages.

Similarly, one may also cite the "meta" package in R using the R code `citation("meta")` as follows:

```
R>citation("meta")
To cite package 'meta' in publications use:
Balduzzi et al. (2019), How to perform a meta-analysis with R: a practical
tutorial, Evidence-Based Mental Health.
```

1.2 Installation of R Software and R Studio

The first step to use R software is to manage the installation of the software. We propose the use of R statistical software because it is an open source, powerful but very flexible software for statistical meta-analysis. One needs to follow these steps in order to download and install R software. Visit the CRAN website in order to download the set-up file: https://cran.r-project.org/bin/windows/base/.

Once the set-up file is downloaded, double click it to begin the installation process. The steps that follow are almost self-explanatory. While installing R, one gets tips at each stage of the installation process. After successfully installing R, one may wish to download R Studio, by visiting the following website: https://www.rstudio.com/home/. For installation to start, one needs to double click on the set-up file that has been downloaded. R is enriched with a number of specific packages that enable one to conduct specific data analysis tasks.

The following steps are followed in order to download and install an R package.

1. Go to the website: http://cran.r-project.org/. Select "Packages."
2. Select "Available CRAN Packages By Name" and choose the package that you wish to install. In the case of meta-analysis, choose "meta," "metafor," or any other package that you may wish to install.
3. The package is then downloaded as a zip file and saved in a folder within your computer.
4. Back in R, install package from a local zip folder.

Alternatively, within R environment, use the command line:
`R>install.packages("package name").`

Finally, it is also possible to use drop down menu in R by selecting, "Tools," then "Packages," and finally "Install packages." Choose to install from Repository CRAN or from a local zip file. You might be prompted to choose the secure CRAN mirrors (listed by country) before you proceed with installation. It is often useful to ensure that the dependencies are set to TRUE. That enables R to directly install all other packages that the package "meta" depends on.

1.3 Basics of Meta-Analysis

Meta-analysis is a statistical method of pooling results from various studies or clinical trials. According to Hoffman (2015), meta-analysis is a set of techniques used to combine the results of a number of different reports, studies, or outcomes into one report, in order to create a single, more precise estimate of an intervention effect. Meta-analysis therefore enables one to come up with a pooled estimate of a particular population parameter based on information from many publications or studies that may have been conducted distinctly from one another. Let us consider the type of variable outcomes in data. The type of outcome determines the types of summary measure and hence the type of data analysis to conduct:

1. Binary Outcome: For instance, consider a status outcome like "dead or alive," "pain free or in pain," "smoking or not smoking." It is only possible that each study participant is in one of two possible mutually exclusive states. These outcomes are dichotomous in nature.
2. Continuous Outcome: Some examples of continuous outcome data include "CD4 cell count of individuals," "viral load," "Body weight," "age" among others. Consider the following question. "What is the viral load after 6 months of Anti Retro-viral Therapy (ART) treatment for patients with HIV?" The answer is possibly a value such as 10,000 copies of HIV RNA in a milliliter of blood. This value could be anything in the range of 0 copies/ml to 100,000 copies/ml (or so), hence it forms a continuous type of outcome.

We take note that meta-Analysis can be performed for various types of Systematic Reviews:

1. Intervention reviews,
2. Diagnostic test accuracy review,
3. Methodology review,
4. Overview of reviews, and
5. Flexible reviews

However, in this chapter, we shall mainly focus on meta-analysis of intervention studies. Such is the case when we have an intervention such as drug treatment whose effect needs to be tested. Two groups of study individuals are then compared, the group exposed to the intervention versus the non-exposed group.

1.4 Effect Measures for Continuous Outcome Data

In order to conduct meta-analysis for intervention studies and when dealing with continuous (or even count) type of outcome data, one needs the following information:

1. Mean of the summary (effect or outcome) measure for the treatment and the control groups
2. Standard deviation of the summary measure for both treatment and control groups.

Therefore, the effect measures that are used for meta-analysis of continuous type of outcome data are thus: Mean Difference (MD), Standardized Mean Difference (SMD) and Ratio of Means (ROM). The mean difference is given by,

$$MD = M_E - M_C$$

while Standardized Mean Difference is measured by dividing the Mean Difference by the standard deviation of the measurements, rather than the standard deviation of the mean difference.

$$SMD = \frac{MD}{SD(measurements)}.$$

Just to mention—without giving the details as at now, some of the methods used to estimate the Standardized Mean Difference include Hedge's g, Cohen's d, and Glass' delta.

1.5 Effect Measures for Dichotomous Outcome Data

Similarly, in order to conduct meta-analysis for intervention studies with dichotomous outcome type of data, one needs the following information:

1. The total number of patients in each treatment group (exposed versus the non-exposed groups, also referred to as the treatment versus control groups).
2. The number of patients who experienced the relevant outcome in each treatment group (treatment and control groups).

The following effect measures are then useful in Meta-Analysis of dichotomous type of outcome: Relative Risk (RR) or Risk Ratio (RR), Odds Ratio (OR), and Risk Difference/Absolute Risk Reduction (ARR).

Let us now consider into detail, the effect measures outlined for meta-analysis of dichotomous outcome from an intervention study.

We first outline a 2X2 contingency table showing incidence of heart attack among individuals who received a stress management intervention, versus those who did not receive stress management intervention. The question is: does stress management for people with heart disease reduce heart attacks? There is a belief that inclusion of stress management in a rehabilitation program may be useful. To study this, a cohort of 328 individuals were included in a Randomized Controlled Trial and the results in Table 1 recorded. For additional information on psychological interventions for coronary heart attacks, you may read Whalley et al. (2011).

Table 1 Stress management for reduction of heart attacks

Groups of patients	Heart attacks	Heart attacks free	Total
Stress Management	a=119	b=45	a+b=164
No stress Management	c=130	d=34	c+d=164
Total	a+c=249	b+d=79	n=328

Computing the Odds Ratio (OR)
Refer to the data in Table 1. The odds of heart attack among the patients who were on stress management (exposed group) is

$$Odds(exposed) = \frac{a}{b} = \frac{119}{45} = 2.64.$$

The odds of heart attack among the patients who never got exposure to stress management is

$$Odds(non - exposed) = \frac{c}{d} = \frac{130}{34} = 3.82.$$

Hence the Odds Ratio (OR) is given by,

$$Odds\ Ratio = \frac{Odds(exposed)}{Odds(non - exposed)} = \frac{ad}{cb}. \tag{1}$$

The confidence interval for the OR is given by,

$$log(OR) \pm Z_{(\alpha/2)} se(log(OR)), \tag{2}$$

where (for a,b,c, and d as illustrated in Table 1)

$$se(log(OR)) = \sqrt{(1/a + 1/b + 1/c + 1/d)}. \tag{3}$$

The following R code can be used to achieve computation of the OR:

Program code: Computation of Odds Ratio

```
install.packages("fmsb")

library(fmsb)

M_OR <- oddsratio(119, 130, 45, 34)

str(M_OR)

print(M_OR)
```

The function "oddsratio" can be used to return the cross tabulation for the data and the computed OR, including its confidence interval. The function returns the analysis output below:

```
List of 5
p.value : num 0.156
conf.int : num [1:2] 0.415 1.152
..- attr(*, "conf.level")= num 0.95
estimate : num 0.692
method : chr "Odds ratio estimate and its significance probability"
data.name: chr "119 130 45 34"
attr(*, "class")= chr "htest"
Odds ratio estimate and its significance probability
data: 119 130 45 34
p-value = 0.1561
95 percent confidence interval:
[0.4152843, 1.1518463]
sample estimates:
[1] 0.6916239
```

An OR of 0.69 implies that stress management reduces the Odds of heart attack by 31%. The confidence interval shows that the effect of stress management varies between a reduction of 58% (lower confidence level = 0.415) to an increase in odds of about 15% (upper confidence level = 1.152)

Computing the Risk Ratio (RR)
We use the data in Table 1. The Risk of heart attack among the patients who were on stress management (exposed group) is

$$Risk(exposed) = \frac{a}{a+b} = \frac{119}{164} = 0.726.$$

The Risk of heart attack among the patients who never got exposure to stress management is

$$Risk(Non-exposed) = \frac{c}{c+d} = \frac{130}{164} = 0.793.$$

Therefore the Risk Ratio is thus,

$$Risk\ Ratio = \frac{Risk(exposed)}{Risk(non-exposed)} = \frac{a(c+d)}{c(a+b)}. \tag{4}$$

The confidence interval for the Risk Ratio is given by,

$$log(RR) \pm Z_{(\alpha/2)} se(log(RR)), \tag{5}$$

where

$$se(log(RR)) = \sqrt{\frac{1}{a} - \frac{1}{a+c} + \frac{1}{c} - \frac{1}{c+d}}. \tag{6}$$

The CI for the RR is obtained by taking exponents of the values obtained from the CI for the log(RR).

An R code for computation of the Risk Ratio and its confidence interval is as follows:

Program code: Computation of Risk Ratio

```
library(fmsb)

RR<-riskratio(119, 130, 164, 164)

str(RR)

print(RR)
```

The R code returns the following output:

```
List of 5
p.value : num 0.156
conf.int : num [1:2] 0.81 1.03
..- attr(*, "conf.level")= num 0.95
estimate : num 0.915
method : chr "Risk ratio estimate and its significance probability"
data.name: chr "119 130 164 164"
- attr(*, "class")= chr "htest"
Risk ratio estimate and its significance probability
data: 119 130 164 164
p-value = 0.1561
95 percent confidence interval:
0.8099204 1.0345819
sample estimates:
[1] 0.9153846
```

One may interpret the RR measure as follows:

1. A Risk Ratio of 0.92 implies that the risk of a heart attack if an individual underwent stress management is reduced by about 8%.
2. The confidence interval shows that the effect of stress management on risk varies between a reduction of 20% (Lower confidence level = 0.8099) to an increase in risk of about 3.4% (Upper confidence level = 1.035).

Absolute Risk Difference (RD)
Risk Difference is defined as

$$RD = Risk(treatment\ group) - Risk(control\ group) \tag{7}$$

$$RD = 0.726 - 0.793 = -0.067$$

and is interpreted as follows: Stress management has the potential to reduce risk of heart attack by about 7%.

The Absolute Risk Difference is given by

$$ARD = |0.726 - 0.793| = 0.067.$$

Additional information on some of these effect measures, also referred to as measures of disease occurrence in most epidemiology text books can be found in Woodward (2013).

2 Methods for Pooling the Effect Measures

Two approaches for pooling effect measures from dichotomous type of outcomes are discussed in detail. Two more are mentioned with little illustration. However, the inverse variance method is the most commonly used approach for meta-analysis due to its ability to handle both continuous and dichotomous kind of outcomes.

- The Peto Odds Ratio
- Mantel Haenszel methods
- Inverse variance method
- O-E and Variance method for combining studies

2.1 The Mantel Haenszel Method

This is a very important tool when we intend to examine the association of a given exposure variable (x) to an outcome variable (y) when there is a third stratification

Table 2 Cross tabulation: Computation of the Mantel Haenszel statistic

k-th study	y=1	y=0	Total
x=1	n_{k11}	n_{k10}	n_{k1}
x=0	n_{k01}	n_{k00}	n_{k10}
Total	m_{k1}	m_{k0}	n_k

variable (z). It becomes important that we do not rely on marginal/crude effect measures but conditional (adjusted) effect measures. We proceed by using the Odds Ratio as the effect measure of interest (Woodward, 2013).

Before we decide to pool the Odds Ratios from different studies, there is a pertinent question to answer. "When do we need to compute a common adjusted odds ratio?" The following options apply whenever we are faced with such a situation.

1. We do not need to pool results when the adjusted effect measures (odds ratios) are significantly different across the different strata. In this case, the stratification variable z is actually an effect modifier. Note that you do not need to pool results that are significantly different.
2. When the adjusted odds ratios are not significantly different, then a common conditional (adjusted) odds ratio can be computed and the Mantel Haenszel estimate of the common OR becomes useful.

Consider the cross tabulation in Table 2 for the k-th study, with the exposure variable (x) denoted as a factor with two levels (0,1) and the outcome variable (y) also denoted as a factor with two levels (0,1). The stratification variable (z) is, however, assumed to have r levels (i.e., $k = 1, 2, \ldots, r$)

The Mantel Haenszel estimate of the common Odds Ratio from the r studies takes the form,

$$\begin{aligned} \widehat{\theta}_{MH} &= \frac{\sum_{k=1}^{r} n_{k11} n_{k00}/n_k}{\sum_{k=1}^{r} n_{k10} n_{k01}/n_k} \\ &= \sum_{k=1}^{r} w_k \widehat{\theta}_k, \end{aligned} \tag{8}$$

where

$$w_k = \frac{n_{k10} n_{k01}/n_k}{\sum_{j=1}^{r} n_{j10} n_{j00}/n_j}. \tag{9}$$

The weight w_k is the inverse variance of θ when $\widehat{\theta}_k$ is near 1.

Example 1: Mantel Haenszel Statistic

In this example, we consider the possible association between HIV status and circumcision from nine National AIDS and STI's Control Programme (NASCOP) regions in Kenya. The data is hypothetical but mimics data from one of the Kenya AIDS Indicator Surveys. It is a cross tabulation of a categorical variable "male

circumcision" and another categorical variable "Human Immuno Deficiency (HIV) Status."

A single cross tabulation of the "male circumcision" versus the "HIV status" for this hypothetical data, for the entire country, is highlighted in the R code for unadjusted OR below. The R package *fmsb* needs to be installed and loaded.

Program code: Unadjusted OR for the hypothetical NASCOP data

```
install.packages("fmsb")

require(fmsb)

oddsratio(123, 70, 4256, 356,conf.level=0.95)
```

The syntax for the function "oddsratio" is as follows; oddsratio (Disease+Exposure, Disease+No-Exposure, Exposure+No-Disease,
No-Disease + No-Exposure, conf.level=0.95).

The R code yields the following outcome:

```
Odds ratio estimate and its significance probability
data: 123, 70, 4256, 356.
p-value < 2.2e-16
95 percent confidence interval: [0.1075070, 0.2009436]
sample estimates: 0.1469791
```

Comment The interpretation of the unadjusted OR is as follows. Male circumcision reduced HIV infection by about 85%. This is the net reduction without considering that HIV prevalence in Kenya may differ from region to region. A pooled result obtained from the weighted ORs for all regions may end up being more informative rather than the absolute OR if the effect of regions was factored in the analysis. The Mantel Haenszel statistic is useful for pooling the results. The R code showing an array with data from each region is given next. The data shows male circumcision versus HIV status by region.

Program code: Data for HIV status stratified by NASCOP regions

```
HIVdata <- array(c(10, 1, 562, 19,

15, 1, 591, 8,
```

```
4, 0, 397, 19,

9, 0, 481, 11,

17, 5, 511, 42,

8, 3, 380, 32,

34, 56, 423, 157,

11, 3, 417, 28,

15, 1, 494, 40),

dim =c(2, 2, 9),

dimnames =list(Circumcision.status =

c("Circumcized", "Uncircumcized"),

HIV.status =c("HIV Infected", "HIV Uninfected"),

NASCOP.region =c("Region 1", "Region 2", "Region 3",

"Region 7", "Region 6", "Region 5", "Region 4",

"Region 8", "Region 9")))

HIVdata
```

The $HIVData$ array contains the crosstabulations for all regions. Results for region 1 and 9 are displayed in Tables 3 and 4.

We first compute the OR for each region. From the summary, we see some differences in the OR from region to region, some regions have high OR (e.g., OR = 0.85) while others have lower (e.g., OR = 0.14). Do we need to pool the results? In later examples, we shall introduce the tests for heterogeneity between studies, using the $Q-statistic$ or similar statistics such as I^2. If there is significant heterogeneity, then one needs to take caution before obtaining pooled estimates of effect measures.

The R code that returns OR for each region reads;

Table 3 NASCOP.region = Region 1

Circumcision.status	HIV Infected	HIV uninfected
Circumcized	10	562
Uncircumcized	1	19

Table 4 NASCOP.region = Region 9

Circumcision.status	HIV Infected	HIV uninfected
Circumcized	15	591
Uncircumcized	1	8

Table 5 OR by "NASCOP Regions"

Regions	Odds ratio	Regions	Odds ratio
Region 1	0.243	Region 2	0.149
Region 3	0.442	Region 4	0.454
Region 5	0.264	Region 6	0.207
Region 7	0.224	Region 8	0.264
Region 9	0.846		

[a] The OR are computed from a cross tabulation of HIV status (positive/negative) versus male circumcision status (yes/no) by region.

Program code: OR by region in the hypothetical NASCOP data

```
library(vcd)

xx<-loddsratio(HIVdata)

yy<-exp(coef(xx));yy
```

The outcome is illustrated in Table 5.

We now compute the classical Mantel Haenszel test using the *mantelhaen.test*() function.

Program code: R code for the Mantel Haenszel test

```
mantelhaen.test(HIVdata)
```

From the results of the Mantel Haenszel test, male circumcision reduces the odds of HIV infection by about 74%. The confidence interval excludes a value 1 and therefore, we are confident that Male Circumcision reduces HIV infection.

```
Mantel-Haenszel chi-squared test with continuity correction
data: HIVdata
Mantel-Haenszel X-squared = 55.371, df = 1, p-value = 9.979e-14
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
[0.1808276, 0.3728469]
sample estimates:
common odds ratio
0.2596556
```

We may also conduct the exact conditional test of independence using the $mantelhaen.test(HIVdata, exact = TRUE)$ function, for which a pooled Odds Ratio of 27.1% is obtained.

Program code: R code for the Mantel Haenszel test

```
mantelhaen.test(HIVdata, exact =TRUE)
```

2.2 The Peto Odds Ratio

This is approximate Odds Ratio and works very well in a case of rare events. While the Mantel Haenszel method can also be used to pool other effect measures such as Risk Ratio, the Peto Odds Ratio can only be used to pool Odds Ratio. It provides an alternative method of pooling OR, to the Mantel Haenszel method described above (Yusuf et al., 1985).

Consider the cross tabulation in Table 6:

For the k-th stratum, we define the following parameters.

$$\begin{aligned} O &= a, \\ E &= ((a+b)(a+c))/n, \\ V &= ((a+b)(c+d)(a+c)(b+d))/(n^2(n-1)) \end{aligned}$$

Table 6 A k-th stratum cross tabulation for computation of the Peto Odds Ratio

	Exposed	Non-Exposed	Total
Cases	a	b	a+b
Non cases	c	d	c+d
Total	a+c	b+d	n

where V is both weighting factor and variance for the difference $(O - E)$. Then, the Peto Odds Ratio is given by,

$$\widehat{\Psi} = \exp\left(\frac{(O-E)}{V}\right). \tag{10}$$

Furthermore, define $n = a + b + c + d$, the total sample for each stratum and $z_{\alpha/2}$, a quantile from the standard normal distribution. Then, the confidence interval for Ψ is given by,

$$CI = \exp\left(\frac{(O-E) \pm z_{\alpha/2}\sqrt{V}}{V}\right). \tag{11}$$

Now, the pooled OR, actually the Peto Odds Ratio, under the assumption of a mixed effects model is given as follows:

$$\widehat{\Psi}_{pool} = \exp\left(\frac{\sum_{k=1}^{r}(O_k - E_k)}{\sum_{k=1}^{r} V_k}\right) \tag{12}$$

and the corresponding Confidence Interval is given by,

$$CI_{pool} = \exp\left(\frac{\sum_{k=1}^{r}(O_k - E_k) \pm z_{\alpha/2}\sqrt{\sum_{k=1}^{r} V_k}}{\sum_{k=1}^{r} V_k}\right). \tag{13}$$

Exercise (Peto Odds Ratio)
The data below represents child mortality after administration of a new vaccine in 5 different studies. The data is provided in columns marked `v_all` for all vaccinated children, `v_cases` for vaccinated and mortality cases, `nv_all` for all non-vaccinated children, `nv_cases` for all non-vaccinated mortality cases. The column marked `study` gives the study identity.

Obtain the OR for each study. Are the ORs significantly different? Is it appropriate to pool the ORs? Obtain the pooled Peto OR. This exercise is meant to be a manual computation exercise. The R codes shall be provided here below for purposes of computing (Table 7).

Table 7 Excercize: Peto Odds Ratio

Study	v_cases	nv_cases	v_all	nv_all
1	50	60	702	535
2	40	45	620	760
3	108	124	710	728
4	28	39	400	325
5	73	62	658	420

An analysis output from R software is provided here below.

Program code: Peto Odds Ratio

```
library(metafor)
column_names<-c("STUDY",    "v_cases", "nv_cases", "v_non cases",
"nv_non_cases", "v_all",    "nv_all")
study1<-c(1,    50,  60,    702-50,  535-60,    702,    535)
study2<-c(2,    40,  45,    620-40,  760-45,    620,    760)
study3<-c(3,    108, 124,   710-108, 728-124, 710,    728)
study4<-c(4,    28,  39,    400-28,  325-39,    400,    325)
study5<-c(5,    73,  62,    658-73,  420-62,    658,    420)
data_vacc<-rbind(study1, study2, study3, study4, study5)
data_vacc<-as.data.frame(data_vacc)
colnames(data_vacc)<-column_names
data_vacc
attach(data_vacc)
rma.peto(ai=v_cases, ci=nv_cases,
bi=v_non_cases, di=nv_non_cases, data=data_vacc)
```

The following results are obtained from the Peto OR analysis.

Fixed-Effects Model (k = 5)
Test for Heterogeneity:
Q = 6.5329, df = 4, p-val = 0.1627
Model Results (log scale):
estimate = -0.2609, se=0.0857, zval=-3.0443, pval=0.0023, ci.lb= -0.4288, and ci.ub=-0.0929
Model Results (OR scale):
estimate =0.7704, ci.lb = 0.6513, ci.ub=0.9113

From the meta-analysis of the log-OR using Peto's method, the Q-statistic is used as a measure of heterogeneity between studies. The `p-value` of $p = 0.163$ indicates that there is no significant heterogeneity in the OR between the studies. We therefore believe that the pooled OR = 0.77 represents the expected situation from the five studies.

On the other hand, the pooled Odds Ratio is significantly different from 1 (confidence interval: [0.65, 0.91]). This implies that the odds of mortality given the child was vaccinated is significantly different from the odds or mortality given the child was not vaccinated. Since the estimated OR is less than 1 ($OR = 0.77$), then based on information from the five studies, the odds of mortality is reduced if the child was vaccinated.

3 Steps of Meta-Analysis for Intervention Studies with R Statistical Software

The steps of meta-analysis involve generating forest plots, funnel plots and furthermore, a discussion on measures of heterogeneity. Using R statistical software, the packages that are useful in meta-analysis include "meta," and "metafor" packages.

The first step is to install the packages using `install.packages("package name")` and this is done only once. However, one needs to load the package using `library(package name)` at every instance of running an analysis or using the meta-analysis functions.

Program code: Meta-analysis packages

```
install.packages("meta", dependencies = TRUE)

install.packages("metafor", dependencies = TRUE)

library(meta)  #Load the library

library(metafor)
```

In this section, we provide the R codes and the outputs from conducting a meta-analysis. Furthermore, a detailed interpretation of results of Meta-Analysis is done. Further information on meta-analysis can be obtained from the textbook by Chen and Peace (2013). The web information by Kapoor and Chetty (2017) is quite useful.

3.1 Fixed Effects Versus Random Effects Meta-Analysis

The philosophy behind fixed effects model is that there is one real value for the treatment effect and that all trials estimate this single value.

On the other hand, the philosophy behind random effects model is that there are many possible real values for the treatment effect (depending on dose, duration, etc.) and that each trial estimates its own real value.

If there is heterogeneity, fixed effect and random effects models may give different pooled estimates and therefore, the interpretations are expected to be different.

Consider this example:

1. A Risk Difference (RD) = 0.3 under fixed effects model means that the best estimate of the one and only one real RD is 0.3.
2. A Risk Difference (RD) = 0.3 under random effects model implies that the best estimate of the mean of all possible real values of the RD is 0.3.
3. Note that the random effects model gives wider confidence interval.
4. In practice, people tend to interpret fixed and random effects the same way. But this should not be the case.

3.2 Some Study Designs in Clinical Trials

It is important to note that meta-analysis of studies cannot be performed without regards to the type of study design that was used in the study. In most literature, the methods for meta-analysis for parallel study designs are illustrated. The knowledge of study designs is important since there are differences in meta-analysis methods when handling studies from different study designs. The Cochran Handbook documents how meta-analysis can be performed for studies from the following study designs (Higgins and Green, 2011).

1. Parallel study design is the most basic design for a meta-analysis of intervention. The design involves grouping study participants into two groups, the experimental group and the control group. Each group receives only one unique treatment during the entire study period (Woodward, 2013).
2. Cross-over study design is designed almost the same way as parallel study designs, except that study participants are switched between the groups at halfway the interval of the study period. Participants who were in the treatment group are switched to the control group and vice versa. The benefit is that each study participant in a treatment group, acts as their own control (Woodward, 2013). This switch has implications on the statistical methods for pooling results from such study designs.
3. Factorial designs is set up in such a way that more than two treatments can be assigned to a small group of study participants in one single experiment. In a

2X2 Factorial design, where we have two treatments A, and B, the participants are cab be split into four groups A+B, B+Control, A+Control, and Control. This is like conducting three experiments in one (Jaki and Vasileiou, 2017).

These study designs can be grouped into two major categories:

1. **Experimental study designs:** Include parallel study designs, dose-ranging designs, cross-over study designs, cluster randomized trials, factorial study designs, step wedge designs, quasi experimental studies, among others.
2. **Observational study designs:** Include cohort studies, cross-sectional studies, case control studies, among others.

For more information on study designs, one may wish to look at some of the following references (Stampfer et al., 1985; Grimmes and Schulz, 2002; Pickering et al., 2019; KC et al., 2019; Donner and Klar, 2000; Hayes et al., 2000; Hemming et al., 2015; Hussey and Hughes, 2007).

3.3 Meta-Analysis of Dichotomous Outcome Data

We shall use a dataset that is available in the meta package to illustrate the steps of meta-analysis for continuous data (Olkin, 1995). The dataset named "Olkin95" summarizes results from 70 articles that studied Thrombolytic Therapy after Acute Myocardial Infarction. The data frame has the following columns:

- author (for first author),
- year (for year of publication),
- event.e (for number of events in experimental group),
- n.e (for number of observations in experimental group),
- event.c (for number of events in control group), and
- n.c (for number of observations in control group).

Here is the R code for meta-analysis of the "Olkin95" data.

Program code: Exploring the "Olkin95" data

```
library(meta); #Load the data

data(Olkin95);

View(Olkin95); #To view the data use the codes

help(Olkin95); #To get help on metabin and data
                functions

try(data(package = "meta"));  #Datasets in meta package
```

Using the "metabin" function in "meta" package, we conduct a meta-analysis for a subset of the Olkin95 data containing only six rows (41, 47, 37, 61, 51, 59). Study 37 specifically has a large sample size. With this large sample size, we expect more precision in estimating the effect measures from study 37. That should reflect in smaller confidence interval for effect measure from such a study when we plot the forest plot.

Program code:"metabin" function for dichotomous data

```
Olkin95_1<-Olkin95[c(41,47,37, 61, 51,59), ]

meta1<-metabin(event.e, n.e, event.c, n.c,

data=Olkin95, subset=c(41,47,37,61,51,59),

sm="RR", method="I")

summary(meta1)
```

Table 8 shows the data from the six studies that have been selected for the subset from the $Olkin95$ dataset.

The effect measure used for the meta-analysis is the Risk Ratio. R does both fixed effect and random effects meta-analysis and returns both results simultaneously. One does not have to specify this assumption during the model specification. A Pooled Risk Ratio = 0.795 is estimated under fixed effects assumption and is very significant ($p\ value < 0.0001$), while the Pooled Risk Ratio = 0.591 is estimated under the random effects assumption and is very significant as well ($p\ value = 0.014$). Measures of heterogeneity including τ^2, $I^2\ and\ Q\ statistic$ shall be discussed in the section on heterogeneity. The meta-analysis tools used include

Table 8 A subset of the "Olkin95" dataset from the R package "meta," used to illustrate meta-analysis of dichotomous type of data

	Author	Year	event.e	n.e	event.c	n.c
41	Schreiber	1986	1	19	3	19
47	Bossaert	1987	4	48	2	39
37	GISSI-1	1986	628	5860	758	5852
61	AIMS	1988	32	502	61	502
51	White	1987	2	107	12	112
59	Meinertz	1988	9	162	19	151

For any intervention study, "event.e" is the number of events in the exposed group and "n.e" is the total number of participants in the exposed group. Whereas "event.c" and "n.c" are similar numbers in the control group.

Inverse Variance method, DerSimonian-Laird estimator for tau^2 and Jackson method for confidence interval of tau^2 and tau.

Some very important figures obtained from meta-analysis include the funnel plot and the forest plot. The R code for generating a funnel plot is shown next:

Program code: R codes for constructing the funnel plot

```
meta1 <-metabin(event.e, n.e, event.c, n.c,

data=Olkin95, subset=c(41,47,37,61,51,59),

studlab = paste(author, year),

sm="RR", method="I")

funnel(meta1)

attach(Olkin95)

funnel(meta1, comb.fixed = TRUE,level = 0.95,

studlab = TRUE, pos.studlab = 4, cex.studlab = 1.25))

### Alternative code for funnel plot

funnel(meta1, comb.fixed = TRUE,

level = 0.95, contour = c(0.9, 0.95, 0.99),

col.contour = c("darkgreen","green","lightgreen"),lwd=2,

cex = 2, pch = 16, studlab = TRUE, cex.studlab = 1.25)

legend(2, 0.2,

c("0.1 > p > 0.05", "0.05 > p > 0.01", "< 0.01"),

fill = c("darkgreen", "green", "lightgreen"))

forest(meta1)
```

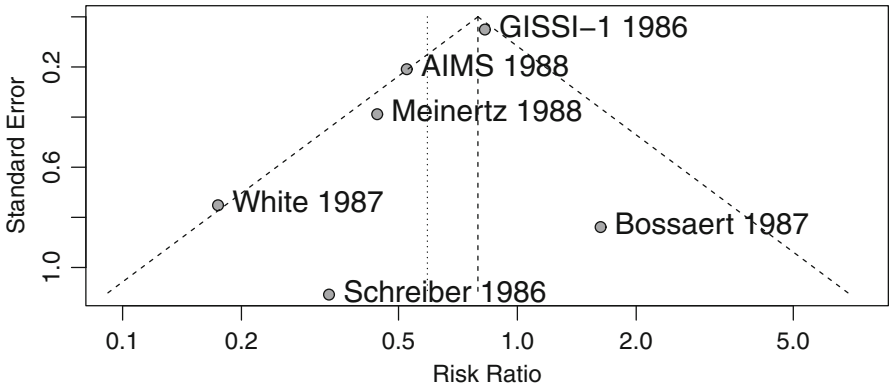


Fig. 1 Funnel plot for a subset data from the "Olkin95" data

Figure 1 shows one of the important outcomes of meta-analysis, the funnel plot. The funnel plots is used to test small study effects or publication heterogeneity so to speak. In the funnel plot from the subset of Olkin95 data used in this study, all dots (representing studies) fall within the triangle hence, all studies contribute fairly well to the meta-analysis. 95% of studies are expected to fall within the triangle. Some of the reasons for asymmetrical funnel plots or small study effects may include (Sterne et al., 2011)

- Poor methodological design or inadequate analysis.
- Reporting bias, including publication bias, location bias, selective outcome and analysis reporting, language bias among others.
- Study Heterogeneity that may result into a correlation between study size and intervention effects.

The forest plot, represented in Fig. 2, is the main output of a meta-analysis. It returns the pooled estimate of the effect measure under both the fixed effects (RR=0.79) and Random effects models (RR=0.59). It also returns the estimates and the confidence intervals of the effect measure for each study, representing them both as quantities and graphically. For instance, study 37: (a study by GISSI-1, 1986) has a very small confidence interval for the Risk Ratio. We see that the study was associated with a very large sample size. Consequently, the study has a weight of 92% and therefore contributes significantly toward the pooled estimate. Compare this with study 51 which contributes only 0.4% weight on the pooled outcome.

The forest plot also returns the measures of heterogeneity. Further interpretation of results of heterogeneity follow in the next section.

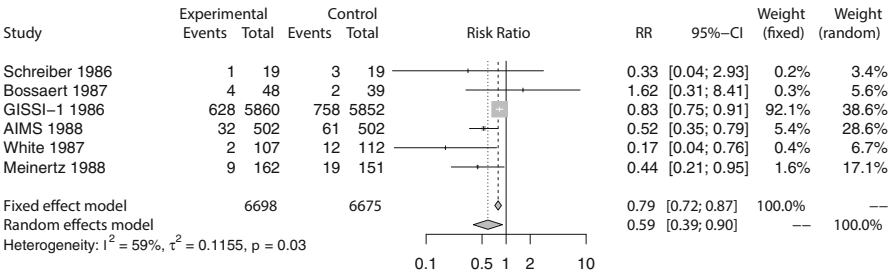


Fig. 2 Forest plot for a subset data from the Olkin95 dataset

3.4 Meta-Analysis of Continuous Outcome Data

Again, we make use of a dataset called "Fleiss93cont" from the "meta" package. To understand the data and the variables in it, use the "help(Fleiss93cont)" function in R.

Program code: The "metacont" function

```
data(Fleiss93cont)

View(Fleiss93cont)

help(Fleiss93cont)

meta_c1 <- metacont(n.e, mean.e, sd.e, n.c, mean.c,sd.c,

data=Fleiss93cont, sm="SMD")

meta_c1

forest(meta_c1)
```

In the meta-analysis (see the options under the "metacont" function), a standardized mean difference is used as the effect measure. The study results show that heterogeneity is very low ($I^2 = 0\%$). We may interpret the estimates from fixed effects model (SMD $= -0.343$). The random effects estimates may be useful when the extent of heterogeneity is assumed to be significant.

We may also wish to allow for other options in the analysis. For instance, let us use Cohen's d instead of Hedges' g, or Glass' delta instead of Hedges' g as method

for estimating Standardized Mean Difference. For an explanation on these methods for estimating SMD, see, for instance (Bakbergenuly et al., 2020)

Program code: Options for metacont function

```
meta5 <- metacont(n.e, mean.e, sd.e, n.c, mean.c, sd.c,
data=Fleiss93cont, sm="SMD", method.smd="Cohen")
meta6 <- metacont(n.e, mean.e, sd.e, n.c, mean.c, sd.c,
data=Fleiss93cont, sm="SMD", method.smd="Glass")
```

4 Heterogeneity

We define heterogeneity as the variation between study results. Among other articles, we reviewed the following and summarized some of the causes of heterogeneity in meta-analysis and how to address heterogeneity (Thompson, 1994; Hardy and Thompson, 1998; Petitti, 2001).

Heterogeneity may occur due to the following reasons:

1. Patients differences including diagnosis, inclusion and exclusion criteria, among others.
2. Intervention differences including the type of intervention, the dose, duration of intervention among others.
3. Outcomes of study. That includes type of outcome, cut-off points, duration of follow-up, among others.
4. Quality and methodology: Either the study is randomized or not, if there is allocation concealment or not, whether blinding is done or not, among other methodology options.

4.1 Handling Heterogeneity Between Studies

What are some of the available options for dealing with high level of heterogeneity between studies?

1. Do not pool studies at all if there is evidence of heterogeneity. One may choose not to meta-analyze since average result may be meaningless in practice.

2. You may ignore heterogeneity and proceed to use a fixed effect model. However, if heterogeneity is very high, interpret fixed effect results with caution.
3. Allow for heterogeneity by using a random effects model.
4. Identify statistical options on how to deal with large variation between studies. This may involve obtaining subgroup estimates, carefully choosing the methods used to estimate the between study variance τ^2, and probably methods for estimating the Standardized Mean Difference.
5. You may also wish to check if the data has any problems such as wrong entries or outright outliers that may influence results.
6. Consider conducting sensitivity analysis. In the context of meta-analysis, this is basically a subgroup analysis that may be based on year of publication, study designs, sample sizes, among other influential covariates.
7. Also investigate clinical and methodological comparability of studies. Sometimes the comparisons being made are not realistic. For instance, one may wish to study the effect of an intervention. But the effect may have been changing over the years. The best practice would be to group studies by periods, rather than lump all studies together.
8. Avoid changing your effect measure without careful consideration of the study designs. The Risk Ratio may not be a good measure for cross-sectional studies.

4.2 Estimation of Study Bias

Two general approaches are used to establish study bias. One approach is to test for funnel plot asymmetry. This can be done by loosely checking if all points plotted with a funnel plot, fall within the triangle of the funnel plot. However, there are also formal tests that can be done to test for funnel plot asymmetry. The other approach is to quantify the extent of between study variation. Let us now consider the two approaches.

Computation of Between-Study Variance The following methods are used to estimate the between-study variance τ^2 :

1. DerSimonian-Laird estimator (In R software, under the "metacont" function, the method.tau = "DL")
2. Paule-Mandel estimator (method.tau = "PM")
3. Restricted maximum-likelihood estimator (method.tau = "REML")
4. Maximum-likelihood estimator (method.tau = "ML")
5. Hunter-Schmidt estimator (method.tau = "HS") among others

Let us Consider the DerSimonian-Laird Estimator Under the random effects model, the assumption of common intervention effect is relaxed and the effect sizes are assumed to have a distribution $\Theta_i = N(\Theta, \tau^2)$, where

$$\tau^2 = max\left\{\frac{Q-(k-1)}{\sum W_i - (\sum W_i^2)/\sum W_i}, 0\right\}.$$

Under the inverse variance method, the weights are given by

$$W_i = \frac{1}{se\{\widehat{\Theta}_i\}^2}.$$

In this case, k is the number of effect measures (or studies) that are being pooled. The Q statistic is the heterogeneity test statistic for inverse variance method for continuous outcomes, but for dichotomous outcomes, either Q_{IV} or Q_{MH} works,

$$Q_{IV} = \sum_{i=1}^{r} w_i(\Theta_i - \widehat{\Theta}_{IV})^2.$$

Test for Funnel Plot Asymmetry By a simple visualization exercise, all points on a funnel plot (study effect measures) are expected to fall within the triangle of the funnel plot. If some studies fall outside the funnel, then they are likely to cause bias to the whole meta-analysis exercise. Such studies should be excluded before a meta-analysis is done.

There are, however, formal tests that can also be done to check for funnel plot asymmetry. The tests for funnel plot asymmetry that have been suggested in the literature include (Sterne et al., 2011):

- Rank correlation test
- Regression test
- Extensions of the regression test

The test, based on a meta-regression by the standard error, aims to detect asymmetry in the funnel plot, which may be an indication of publication bias. However, results should be taken with caution, especially if any of the following situations hold. That there are too few studies (at least 10 studies were suggested by Sterne et al. (2011), if the sample sizes are too similar or if there are outliers or influential studies or subgroups. The test for Funnel Plot Symmetry can be done using the "metabias()" function.

Program code: The "metabias" function

```
data(Olkin95)

meta1 <- metabin(event.e, n.e, event.c, n.c,
```

```
                data=Olkin95, subset=1:10,

                sm="RR", method="I")

metabias(meta1)

metabias(meta1, plotit=TRUE)
```

4.3 A Discussion on Measures of Heterogeneity

Consider the output to the "metacont()" function applied to the "Fleiss93cont" data, for a discussion on the measures of heterogeneity.

Program code: heterogeneity in meta-analysis

```
library(meta)

data(Fleiss93cont)

meta1 <- metacont(n.e, mean.e, sd.e, n.c, mean.c, sd.c,

                  data = Fleiss93cont, sm = "MD")

help(Fleiss93cont)

summary(meta1)

forest(meta1)
```

Confidence Interval Lines on Forest Plot Check how the Confidence Interval lines on the forest plot overlap. If the lines do not overlap, then there are signs of heterogeneity. In this case, the confidence interval lines overlap and thus heterogeneity is not high.

The Chi2 Test The test measures the amount of variation between studies, and tells us if it is more than would be expected by chance. Small p values suggest that heterogeneity is present, the null hypothesis being that there is no heterogeneity.

This test is, however, not very good at detecting heterogeneity. Often a cut-off of $p < 0.10$ is used, but lack of statistical significance may not directly imply there is no heterogeneity.

The I^2 Statistic The I^2 is the proportion of variation that is due to heterogeneity rather than chance. Large values of I^2 suggest heterogeneity. Roughly speaking, I^2 values of 25%, 50%, and 75% could be interpreted as indicating low, moderate, and high heterogeneity (Higgins et al., 2003). In the example above, $I^2 = 29.3\%$ which can be classified as fairly low heterogeneity.

The statistic is computed as below. Define the effect measures by $\widehat{\Theta} = ln(OR)$. Of course, other possible effect measures include:

1. $\widehat{\Theta}$ = ln(RR),
2. $\widehat{\Theta}$ = RD (Risk Difference),
3. $\widehat{\Theta}$ = MD (Mean Difference).

Based on the Mantel Haenszel pooled effect measure, and assuming the null hypothesis of no differences is true, the heterogeneity statistic (for r studies; $k = 1, 2, ..., r$) is given by,

$$Q_{MH} = \sum_{i=1}^{r} w_i(\Theta_i - \widehat{\Theta}_{MH})^2 \sim \chi_{(r-1)|H_0}.$$

The weights are calculated as, $w_i = 1/se(\widehat{\Theta}_i)^2$ and not the weights that were used in the Mantel Haenszel calculation. The I^2 statistic is given by,

$$I^2 = max\left\{100\left(\frac{Q_{MH} - (k-1)}{Q_{MH}}\right), 0\right\}.$$

The same can be done if the pooling method used was the Inverse variance method, or the $O - E$ and variance method, by just replacing the Q_{MH} statistic, and the $se(\widehat{\Theta})$ with the corresponding one.

Tau Squared (τ^2) for Random Effects Models In random effects meta-analysis, the extent of variation on the effects observed from different studies (between-study variance) is referred to as tau squared, τ^2, or Tau2 (Deeks et al., 2008). τ^2 is the variance of the effect size parameters across the studies and it reflects the variance of the true effect sizes. The larger the value of τ^2 or τ, the larger the amount of true heterogeneity (Borenstein et al., 2009). To discuss heterogeneity, the value of τ^2 may be compared to another estimate or a self-defined threshold, in order to comment on how large or small the value is. For random effects models, it is assumed that the common intervention effect is not true and the distribution of effect sizes is assumed to be,

$$\widehat{\Theta} \sim N(0, \tau^2),$$

where τ^2 is estimated by

$$\tau^2 = max\left\{\frac{Q-(k-1)}{\sum w_i - (\sum w_i^2)/\sum w_i}, 0\right\}$$

and $w_i = 1/se(\widehat{\Theta}_i)^2$. The Q statistic can be replaced by the heterogeneity test statistic for inverse variance method,

$$Q_{IV} = \sum_{i=1}^{r} w_i(\Theta_i - \widehat{\Theta}_{IV})^2$$

for continuous outcomes, but for dichotomous outcomes, either Q_{IV} or Q_{MH} can still be used.

4.4 Subgroup Analysis or Sensitivity Analysis

Sensitivity analysis in meta-analysis involves estimating the net effect of a single or group of studies on the pooled result. The approach is to remove such study or studies from the meta-analysis and check for differences in the pooled estimates or shift in extent of heterogeneity that has been caused. This is essentially the outcome of subgroup analysis.

We revert back to the "Olkin95" data and conduct a subgroup analysis. One example of interest would be to conduct a subgroup analysis of studies done before 1973 in the Olkin95 data. The following R code would be useful.

Program code: Subgroup analysis

```
library(meta)

meta2a <- metabin(event.e, n.e, event.c, n.c,

     data=Olkin95, subset=Olkin95$year<1973,

     sm="RR", method="I")

summary(meta2a)

forest(meta2a)
```

Table 9 Results on pooled Risk Ratio from subgroup analysis

Result	Pre 1973 value	Pre 1973 CI	Entire data value	Entire data CI
Fixed effect RR	0.83	[0.68, 1.02]	0.78	[0.74, 0.82]
Random effect RR	0.88	[0.67, 1.16]	0.77	[0.71, 0.83]
I^2	25%	[0%, 66.9%]	17.1%	[0%, 39%]
τ^2	0.03	[0, 0.84]	0.012	[0, 0.15]
τ	0.18	[0, 0.92]	0.11	[0, 0.38]

The results for studies conducted before 1973 and also for the entire Olkin95 dataset are shown in Table 9.

The studies conducted before 1973 showed no significant increase in risk of Acute Myocardial Infarction in patients who were exposed to thrombolytic therapy. For instance, the confidence intervals contain 1 (pooled Fixed Effect risk ratio = 0.83 [0.68, 1.02]). The risk was, however, significantly different when analysis was done for all the 70 studies (pooled Fixed Effect risk ratio = 0.78 [0.74, 0.82]). On heterogeneity, the lower the I^2 value, the less the extent of heterogeneity between the studies. We also see that the extent of heterogeneity was lower for all studies ($I^2 = 17.1\%[0\%, 39\%]$) as opposed to pre-1973 studies ($I^2 = 25\%[0\%, 66.9\%]$). To some extent, the subgroup analysis yields totally different results of the net effect on Acute Myocardial Infarction in patients who were exposed to thrombolytic therapy.

A very good illustration of subgroup analysis can also be found in Macdonald et al. (2020), who examined the effectiveness of primary care interventions against physical frailty among community-dwelling older adults aged over 60 years. The studies that were included in the meta-analysis addressed various forms of primary care including nutrition supplementation, education and comprehensive geriatric assessment or a combination of the interventions. Results of this study by Macdonald et al. (2020) showed significant differences between the subgroups. They reported the following examples of differences between subgroups:

- Interventions using predominantly resistance-based exercise and nutrition supplementation improved frailty status versus control (RR = 0.62 (CI 0.48–0.79), $I^2 = 0\%$).
- Exercise plus nutrition education reduced frailty (RR = 0.69 (CI 0.58–0.82), $I^2 = 0\%$).
- Exercise alone appeared superior to control in improving gait speed (SMD = 0.36 (CI 0.10–0.61, $I^2 = 74\%$), leg strength (SMD = 0.61 (CI 0.09–1.13), $I^2 = 87\%$), and grip strength (Mean Difference = 1.08 (CI 0.02–2.15), $I^2 = 71\%$). In this case, a high degree of heterogeneity was observed.

5 Further Topics in Meta-Analysis

5.1 Meta-Regression

Just like in primary subject level studies, one may conduct a multiple regression for a set of covariates on an outcome variable of interest. For Meta-analysis, in which the raw data is a summary of study level information, meta-regression can be done in much the same way as multiple regression is done. The two approaches adopt the same principles.

In this illustration, we make use of "Fleiss93cont" dataset in the R package "meta." In order to perform regression, we add some (fictitious) grouping variables or covariates to the "Fleiss93cont" data. Suppose one of the variables added is "age" which represents the average age of participants included in each of the included studies, while "region" represents the place where the raw data for each study was collected. The assumption is that these two covariates may explain the effect size within each study.

Program code: Meta-regression

```
library(meta)

data(Fleiss93cont)

View(Fleiss93cont)

Fleiss93cont$age <- c(55, 65, 55, 65, 55)

Fleiss93cont$region <- c("Europe", "Europe", "Asia",

"Asia","Europe")

head(Fleiss93cont)

meta1 <- metacont(n.e, mean.e, sd.e, n.c, mean.c, sd.c,

                  data = Fleiss93cont, sm = "MD")

help(Fleiss93cont)

summary(meta1)

#Conducting meta regression using the update function
```

```
on the metacont function

mu1 <- update(meta1, byvar = region)

summary(mu1)
```

We run a model using the "metacont()" function in the meta package (see the R code on meta-regression). The fixed effects Mean Difference (Pooled MD fixed) is -0.7094 on a confidence interval $[-1.2585; -0.1603]$. The confidence interval gives an indication that the five studies ($k = 5$) included in the meta-analysis return a pooled mean difference that is significantly different from zero. The same interpretation is seen for the Random effects model, where Pooled MD random $= -0.7373[-1.4577; -0.0170]$.

The studies are therefore not quite heterogeneous, with $I^2 = 29.3\%[0.0\%; 72.6\%]$. However, the upper confidence interval is quite large. At 72.6%, we could infer high level of heterogeneity between the studies. If heterogeneity is high, one of the analysis objectives could be to understand some of the covariates that could be possible determinants of the outcome of interest within each study. This could be one of the reasons for conducting meta-regression.

We begin by trying to explore if region had any effect on the Mean Difference reported for each of the pooled studies. To test the null hypothesis that a coefficient is equal to zero, $Z-$ test is often used in meta-analysis (instead of the $t - test$ as used in multiple regression). For two or more covariates, the $Q-$ test is useful. Note that for a single covariate, the result of the $Q-$ test is similar to those of the $Z-$ test.

One can improve the above model by incorporating options such as "tau.common," which is a logical option indicating whether tau squared (τ^2) should be the same across subgroups. Another option is to specify whether we wish to conduct a random effects or fixed effects meta-analysis. The option "comb.fixed" is a logical option, such that if TRUE, then a fixed effects meta-analysis should be conducted. A similar option "comb.random" is also a logical function, such that if TRUE, then a random effects meta-analysis should be conducted.

Program code: Meta-regression with additional options

```
mu2 <- update(meta1, byvar = region, tau.common = TRUE,

comb.fixed = FALSE)
```

5.2 Network Meta-Analysis

This chapter has mainly addressed meta-analysis in the context of a single intervention. For instance, we have looked at the meta-analysis involving an intervention such as use of drug A versus control experiment. An effect measure such as Odds Ratio that was reported in each study is collected and consequently used in the meta-analysis.

In Network Meta-Analysis (NMA), also referred to as multiple treatment meta-analysis, or mixed treatment comparison, the aim is to synthesize the effect sizes of several studies in the presence of multiple interventions or treatments (Shim et al., 2019). For example, a study compared the effects of six anti-hypertensive treatments (including beta-blocker, Angiotensin receptor blockers (ARB), Angiotensin-Converting Enzyme Inhibitors (ACE inhibitors), Diuretic, calcium channel blocker (CCB) and a placebo) for hypertension on the incidence of diabetes (Neupane et al., 2014). In this study, the R codes used for Network Meta-analysis included R "gemtc," "pcnetmeta," and "netmeta," all which are freely available software tools implemented in R.

Acknowledgments The R package "meta" and the help resources (data and functions) from the meta package have largely been used to develop information on the meta-analysis section. You may also refer to the "meta" package manual for additional information on conducting meta-analysis as illustrated in this chapter.

This work was supported through the DELTAS Africa Initiative Grant No. 107754/Z/15/Z-DELTAS Africa SSACAB. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

References

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence-based mental health, 22*(4), 153–160.

Bakbergenuly, I., Hoaglin, D. C., & Kulinskaya, E. (2020). Estimation in meta-analyses of mean difference and standardized mean difference. *Statistics in Medicine, 39*(2), 171–191.

Borenstein M., Hedges L.V., Higgins J.P.T, & Rothstein H.R. (2009). *Introduction to Meta-Analysis*. London: Wiley.

Chen, D. G. D., & Peace, K. E. (2013). *Applied meta-analysis with R*. New York: CRC Press.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Chapter 9: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): Wiley.

Donner, A., & Klar, N. (2000). *Design and Analysis of Cluster Randomised Trials in Health Research* (pp. 131). London: Arnold.

Grimes, D. A., & Schulz, K. F. (2002). An overview of clinical research: the lay of the land. *The Lancet, 359*(9300), 57–61.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine, 17*(8), 841–856.

Hayes, R. J., Alexander, N. D., Bennett, S., & Cousens, S. N. (2000). Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research, 9*(2), 95–116.

Hemming, K., Lilford, R. J., & Girling, A. J. (2015). Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Statistics in Medicine, 34*, 181—196.

Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for Systematic Reviews of interventions* (Vol. 4). New York: Wiley.

Higgins, J. P. T. et al. (2003). Measuring inconsistency in meta- analyses. *BMJ, 327*, 557–560.

Hoffman, J. I. (2015). *Biostatistics for medical and biomedical practitioners*. New York: Academic Press.

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*, 182—91.

Jaki, T., & Vasileiou, D. (2017). Factorial versus multi-arm multi-stage designs for clinical trials with multiple treatments. *Statistics in Medicine, 36*(4), 563–580.

Kapoor, Y., & Chetty, P. (2017). In *Interpretation of results of meta-analysis on different types of plot*. (Available via Project GURU). https://www.projectguru.in/publications/interpretation-meta-analysis-different-plot/.

KC, A., Ewald, U., Basnet, O., Gurung, A., Pyakuryal, S. N., Jha, B. K., Bergström, A., Eriksson, L., Paudel, P., Karki, S., & Gajurel, S. (2019). Effect of a scaled-up neonatal resuscitation quality improvement package on intrapartum-related mortality in Nepal: a stepped-wedge cluster randomized controlled trial. *PLoS Medicine, 16*(9), e1002900.
and Gajurel, S., 2019

Macdonald, S. H., Travers, J., Shé, É. N., Bailey, J., Romero-Ortuno, R., Keyes, M., O'Shea, D., & Cooney, M. T. (2020). Primary care interventions to address physical frailty among community-dwelling adults aged 60 years or older: A meta-analysis. *PloS One, 15*(2), e0228821. https://doi.org/10.1371/journal.pone.0228821.

Neupane, B., Richer, D., Bonner, A. J., Kibret, T., & Beyene, J. (2014). Network meta-analysis using R: a review of currently available automated packages. *PloS One, 9*(12), e115065. https://doi.org/10.1371/journal.pone.0115065.

Olkin, I. (1995). Statistical and theoretical considerations in meta-analysis. *Journal of Clinical Epidemiology, 48*, 133–146.

Petitti, D. B. (2001). Approaches to heterogeneity in meta-analysis. *Statistics in Medicine, 20*(23), 3625–3633.

Pickering, A. J., Njenga, S. M., Steinbaum, L., Swarthout, J., Lin, A., Arnold, B. F., Stewart, C.P., Dentz, H.N., Mureithi, M., Chieng, B., & Wolfe, M. (2019). Effects of single and integrated water, sanitation, handwashing, and nutrition interventions on child. soil-transmitted helminth and Giardia infections: A cluster-randomized controlled trial in rural Kenya. *PLoS Medicine, 16*(6), e1002841.
and Wolfe, M.,

Shim, S. R., Kim, S. J., Lee, J., & Rücker, G. (2019). Network meta-analysis: application and practice using R software. *Epidemiology and Health, 41*, e2019013. https://doi.org/10.4178/epih.e2019013.

Stampfer, M. J., Buring, J. E., Willett, W., Rosner, B., Eberlein, K., & Hennekens, C. H. (1985). The 2× 2 factorial design: Its application to a randomized trial of aspirin and US physicians. *Statistics in Medicine, 4*(2), 111–116.

Sterne et al. (2011). Research Methods and Reporting Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, 1–8. https://doi.org/10.1136/bmj.d4002

Thompson, S. G. (1994). Systematic Review: Why sources of heterogeneity in meta-analysis should be investigated. *Bmj, 309*(6965), 1351–1355.

Whalley, B., Rees, K., Davies, P., Bennett, P., Ebrahim, S., Liu, Z., . . . & Taylor, R. S. (2011). Psychological interventions for coronary heart disease. *Cochrane Database of Systematic Reviews, 8*, CD002902. https://doi.org/10.1002/14651858.CD002902.pub3. Accessed 13 September 2022.

Woodward, M. (2013). *Epidemiology: Study design and data analysis*. London: Chapman and Hall/CRC.

Yusuf, S., Peto, R., Lewis, J., Collins, R., & Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases, 27*(5), 335–371. https://doi.org/10.1016/s0033-0620(85)80003-7. PMID: 2858114.

Longitudinal Meta-Analysis of Multiple Effect Sizes

Alfred Musekiwa, Samuel O. M. Manda, Henry G. Mwambi, Ding-Geng (Din) Chen, Samuel A. Abariga, Michael McCaul, Eleanor Ochodo, and Anke Rohwer

Abstract Meta-analysis methods for univariate effect sizes are well-known and developed. However, multiple outcomes are increasingly being measured and reported in medical research studies, which may lead to multiple effect sizes being estimated. The estimated effect sizes could be correlated because they are measured from the same studies. Additionally, the outcomes are often measured longitudinally, resulting in multiple effect sizes estimated repeatedly over time. Thus, the estimated effect sizes could be correlated within studies both cross-sectionally and serially due to the repeated estimation of the same effect over time in the same study. This results into longitudinal multiple effect sizes. This chapter proposes methods for statistical meta-analysis combining summary data from more than one longitudinal study with multiple effect sizes. The proposed methods are illustrated by an analysis of an example involving longitudinal meta-analysis of HIV studies assessing the effect of some antiretroviral drugs in improving viral load suppression and increasing CD4 count at weeks 4, 8, 12, 16, 20, 24, 32, 40, and 48 after start of treatment assignment.

A. Musekiwa (✉)
School of Health Systems and Public Health, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
e-mail: Alfred.Musekiwa@up.ac.za

S. O. M. Manda
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa
e-mail: Samuel.Manda@up.ac.za

H. G. Mwambi
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
e-mail: mwambih@ukzn.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_8

Keywords Longitudinal meta-analysis · Univariate meta-analysis · Multivariate meta-analysis · Multiple outcomes · Multivariate effect sizes · Time points · Longitudinal · Longitudinal multiple effect sizes · Meta-analysis model · Continuous outcome · Covariate · Random effects · Fixed effects · Residuals · General linear mixed model · Joint distribution · Design matrix · Multivariate normal distribution · Variance–covariance matrix · Covariance · Maximum likelihood · Restricted maximum likelihood · Covariance structures · Random time effects · Within-study effect sizes · Independent effect sizes · HIV studies · Antiretroviral drugs · Viral load suppression · CD4 count

1 Introduction

Univariate meta-analysis combines individual effect sizes such as risk ratios from several studies to obtain an overall effect size. The methods for univariate meta-analysis are well-known (Hedges and Olkin, 1985, Egger et al., 2008, Sutton et al., 2000, Lipsey & Wilson, 2001, Whitehead, 2002, Litell et al., 2008, Higgins et al., 2008, Cooper et al., 2009, Borenstein et al., 2009, Pigott, 2012), and it can be implemented in standard statistical software such as using STATA command metan (Bradburn et al., 1998), metafor package in R (Viechtbauer, 2010), and the mixed procedure in SAS (SAS Institute, 2013). There are also common routine

D.-G. (Din) Chen
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa
e-mail: din.chen@up.ac.za

S. A. Abariga
Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
e-mail: Samuel.Abariga@downstate.edu

M. McCaul
Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

South African GRADE Network, Stellenbosch, South Africa
e-mail: mmccaul@sun.ac.za

E. Ochodo
South African GRADE Network, Stellenbosch, South Africa

Centre for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

A. Rohwer
Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
e-mail: arohwer@sun.ac.za

computer packages that can perform univariate meta-analysis such as MetaWin (Rosenberg et al., 2000), WEasyMA (Chevarier et al., 2000), Review Manager (Review Manager, 2014), MIX (Bax et al., 2006), Comprehensive Meta-analysis (Borenstein et al., 2005), and OpenMetaAnalyst (Wallace et al., 2009, 2012). However, most biomedical studies measure multiple outcomes; for instance, HIV studies investigating the effectiveness of antiretroviral drugs among HIV-infected individuals usually measure both CD4 count and viral load after treatment. The multiple outcomes measured from the same study are usually correlated; for instance, the CD4 count and viral load are known to be (negatively) correlated. A meta-analysis of these correlated multiple effect sizes can take two forms: (1) univariate meta-analyses involving separate independent meta-analyses for each effect size, which ignores correlation between effect sizes, or (2) multivariate meta-analysis where multiple effect sizes are jointly synthesized while taking account of the correlation between them (Hedges and Olkin, 1985; Rosenthal and Rubin, 1986; Raudenbush et al., 1988; Gleser and Olkin, 1994; Berkey et al., 1996; Kalaian and Raudenbush, 1996; Berkey et al., 1998; Van Houwelingen et al., 2002; Nam et al., 2003; Arends et al., 2003; Riley, 2009; Jackson et al., 2011; Mavridis and Salanti, 2012). Although parameter estimates from independent univariate meta-analyses and multivariate meta-analysis are usually similar, multivariate meta-analysis yields more precise estimates of overall effect sizes (Riley, 2009).

Longitudinal studies, which report outcomes at fixed predetermined time points, are increasingly being used in medical research (Verbeke et al., 2014). As a result, the meta-analysis of these longitudinal studies has also become a priority when analyzing the effect of exposure or treatment across a long period of study. A longitudinal meta-analysis model utilizing the general linear mixed model was first proposed by Ishak et al. (2007), and we have recently applied random effects models using both study and random time effects to a meta-analysis of 17 longitudinal randomized controlled trials comparing radiation therapy with and without adjuvant chemotherapy for the postoperative treatment of malignant gliomas, where survival was measured at 6, 12, 18, and 24 months post-randomization (Musekiwa et al., 2016) using data from Fine et al. (1993).

Furthermore, longitudinal studies can also measure multiple effect sizes at each time point leading to longitudinal multiple effect sizes. For example, CD4 count and viral load can be both measured at four-weekly intervals after antiretroviral treatment of HIV-infected individuals in order to assess the impact of the treatment over time. A meta-analysis of longitudinal multiple effect sizes needs to take account of the correlation between the effect sizes, which exists both between different effect sizes and serially over repeated measurements of the same effect size. In this chapter, statistical methods for the meta-analysis of longitudinal studies with multiple effect sizes are proposed. The methods are illustrated by an analysis of an example involving longitudinal meta-analysis of HIV studies assessing the effect of some antiretroviral drugs in improving viral load suppression and increasing CD4 count at weeks 4, 8, 12, 16, 20, 24, 32, 40, and 48 after start of treatment assignment. Section 2 covers the statistical model for longitudinal meta-analysis with multiple outcomes, estimation methods are given in Sect. 3, four models with

different covariance structures are given in Sect. 4 with a practical example for the application of these models in Sect. 5, and the summary of the chapter is given in Sect. 6.

2 Statistical Meta-Analysis Model

We adapt the notation of a linear mixed model from West et al. (2014). Let Y_{ilt} represent the measure of a continuous response variable Y taken from the i-th study ($i = 1, \cdots, n$) (unit of analysis), for the l-th outcome ($l = 1, \cdots, m$), and t-th time point ($t = 1, \cdots, T_i$). The variable Y represents estimated effect size such as the log odds ratio for a binary outcome or a (standardized or non-standardized) mean difference for a continuous outcome. The T_i longitudinal observations for study i allow variable time points per study. A linear mixed model for Y_{ilt} is represented by Eq. (1),

$$Y_{ilt} = \beta_{l1}X_{ilt}^{(1)} + \beta_{l2}X_{ilt}^{(2)} + \cdots + \beta_{lp}X_{ilt}^{(p)} + \delta_{il1}Z_{ilt}^{(1)} + \delta_{il2}Z_{ilt}^{(2)} + \cdots + \delta_{ilq}Z_{ilt}^{(q)} + e_{ilt}, \tag{1}$$

where $X^{(1)}, \cdots, X^{(p)}$ represent the set of p covariates associated with the fixed effects $\beta_{l1}, \cdots, \beta_{lp}$ for the l-th outcome ($l = 1, \cdots, m$); for simplicity, the model does not have an intercept term. For $k = 1, \cdots, p$, $X_{ilt}^{(k)}$ represent the observed value of the covariate $X^{(k)}$ for the i-th study, l-th outcome at t-th time point. Similarly, $Z^{(1)}, \cdots, Z^{(q)}$ is the set of q covariates associated with the random effects $\delta_{il1}, \cdots, \delta_{ilq}$ that are specific to study i and l-th outcome. Both the X's and Z's covariates are allowed to be either continuous or indicator variables. We assume that the X's can be either time-invariant characteristics of the individual study (e.g., study quality) or time-varying for each measurement (e.g., time point of measurement). The last term e_{ilt} is the residual associated with Y_{ilt}. The random effects and residuals in Eq. (1) are random variables, with values drawn from distributions defined below. For a given study, the residuals are assumed to be independent of the random effects.

We now present the model using matrix notation. Let $\boldsymbol{Y}_i$ represent the stacked vector of all estimated effect sizes from the ith study, that is,

$$\boldsymbol{Y}_i = (Y_{i11}, \cdots, Y_{i1T_i} | \cdots | Y_{im1}, \cdots, Y_{imT_i})', \tag{2}$$

where, for simplicity, we assume that m outcomes are observed at each of the T_i time points for study i. We note that the number of time points in $\boldsymbol{Y}_i$ may vary from study to study. Now, by extending Eq. (1), we can model the longitudinal meta-analysis of n studies with m multiple outcomes at each of T_i time points using a general linear mixed model (Laird and Ware, 1982),

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{e}_i, \tag{3}$$

where $\boldsymbol{X}_i$ is an $mT_i \times mp$ design matrix of mp fixed effects to be estimated,

$$\boldsymbol{\beta} = (\beta_{11}, \cdots, \beta_{1p} | \cdots | \beta_{m1}, \cdots, \beta_{mp})'.$$

Like $\boldsymbol{X}_i$, $\boldsymbol{Z}_i (\subseteq \boldsymbol{X}_i)$ is an $mT_i \times mq$ design matrix of mq random effects,

$$\boldsymbol{\delta}_i = (\delta_{i11}, \cdots, \delta_{i1q} | \cdots | \delta_{im1}, \cdots, \delta_{imq})'.$$

The residuals for $\boldsymbol{Y}_i$ are contained in the $mT_i \times 1$ vector

$$\boldsymbol{e}_i = (e_{i11}, \cdots, e_{i1T_i} | \cdots | e_{im1}, \cdots, e_{imT_i})'.$$

For each study, the residuals and random effects are assumed to be independent both individually and between each other, $\text{cov}(\boldsymbol{\delta}_i, \boldsymbol{e}_i) = 0$.

We also assume, without loss of generality, that the joint distribution of random effects is $\mathbf{0}$-centered $\boldsymbol{\delta}_i \sim MVN(\mathbf{0}, \boldsymbol{D})$ (multivariate normal distribution), where $\boldsymbol{D}$ is an $mq \times mq$ variance–covariance matrix consisting of symmetric matrices in the diagonals and non-symmetric off-diagonals,

$$\begin{pmatrix}
\begin{bmatrix} \tau_{11}^2 & \tau_{11,12} & \cdots & \tau_{11,1q} \\ & \tau_{12}^2 & \cdots & \tau_{12,1q} \\ & & \ddots & \vdots \\ & & & \tau_{1q}^2 \end{bmatrix} & \begin{bmatrix} \tau_{11,21} & \tau_{11,22} & \cdots & \tau_{11,2q} \\ \tau_{12,21} & \tau_{12,22} & \cdots & \tau_{12,2q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1q,21} & \tau_{1q,22} & \cdots & \tau_{1q,2q} \end{bmatrix} & \cdots & \begin{bmatrix} \tau_{11,m1} & \tau_{11,m2} & \cdots & \tau_{11,mq} \\ \tau_{12,m1} & \tau_{12,m2} & \cdots & \tau_{12,mq} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1q,m1} & \tau_{1q,m2} & \cdots & \tau_{1q,mq} \end{bmatrix} \\
\begin{bmatrix} \tau_{21,11} & \tau_{21,12} & \cdots & \tau_{21,1q} \\ \tau_{22,11} & \tau_{22,12} & \cdots & \tau_{22,1q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{2q,11} & \tau_{2q,12} & \cdots & \tau_{2q,1q} \end{bmatrix} & \begin{bmatrix} \tau_{21}^2 & \tau_{21,22} & \cdots & \tau_{21,2q} \\ & \tau_{22}^2 & \cdots & \tau_{22,2q} \\ & & \ddots & \vdots \\ & & & \tau_{2q}^2 \end{bmatrix} & \cdots & \begin{bmatrix} \tau_{21,m1} & \tau_{21,m2} & \cdots & \tau_{21,mq} \\ \tau_{22,m1} & \tau_{22,m2} & \cdots & \tau_{22,mq} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{2q,m1} & \tau_{2q,m2} & \cdots & \tau_{2q,mq} \end{bmatrix} \\
\vdots & \vdots & \ddots & \vdots \\
\begin{bmatrix} \tau_{m1,11} & \tau_{m1,12} & \cdots & \tau_{m1,1q} \\ \tau_{m2,11} & \tau_{m2,12} & \cdots & \tau_{m2,1q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{mq,11} & \tau_{mq,12} & \cdots & \tau_{mq,1q} \end{bmatrix} & \begin{bmatrix} \tau_{m1,21} & \tau_{m1,22} & \cdots & \tau_{m1,2q} \\ \tau_{m2,21} & \tau_{m2,22} & \cdots & \tau_{m2,2q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{mq,21} & \tau_{mq,22} & \cdots & \tau_{mq,2q} \end{bmatrix} & \cdots & \begin{bmatrix} \tau_{m1}^2 & \tau_{m1,m2} & \cdots & \tau_{m1,mq} \\ & \tau_{m2}^2 & \cdots & \tau_{m2,mq} \\ & & \ddots & \vdots \\ & & & \tau_{mq}^2 \end{bmatrix}
\end{pmatrix},$$

where $\tau_{lj,l'j'} = \text{cov}(\delta_{ilj}, \delta_{il'j'})$, the covariance between the j-th and j'-th random effects for effect sizes l and l', respectively.

Similarly, the joint distribution of residuals is assumed $\mathbf{0}$-centered $\boldsymbol{e}_i \sim MVN(\mathbf{0}, \boldsymbol{S}_i)$ with $mT_i \times mT_i$ variance–covariance matrix $\boldsymbol{S}_i$ similar to $\boldsymbol{D}$ above except that τ is now replaced by σ_i. Therefore, in $\boldsymbol{S}_i$, we have $\sigma_{ilt,il't'} = \text{cov}(e_{ilt}, e_{il't'}), \forall t \neq t', l \neq l'$, the covariance between the t-th and t'-th residual terms for effect sizes l and l', respectively.

Marginally, $\boldsymbol{Y}_i \sim MVN(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_i)$, where $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i{}' + \boldsymbol{S}_i$. The within-study and between-study correlations between effect sizes are determined by the covariance structures imposed on $\boldsymbol{S}_i$ and $\boldsymbol{D}$, respectively. We discuss the various covariance structures in Sect. 4. The goal of meta-analysis is to estimate the

parameters in the vector $\boldsymbol{\beta}$ and the variance and covariance parameters in $\boldsymbol{D}$ and $\boldsymbol{S}_i$.

3 Estimation of Parameters

The maximum likelihood (ML) and restricted maximum likelihood (REML) are the two most commonly used methods of estimation, and we discuss them in this section.

3.1 Maximum Likelihood (ML) Estimation

Let $\boldsymbol{\alpha}$ denote the vector of all variance and covariance parameters found in $\boldsymbol{V}_i(\boldsymbol{\alpha}) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i{}' + \boldsymbol{S}_i$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha})'$ be the $s-$dimensional vector of all parameters in the marginal model for $\boldsymbol{Y}_i$. The marginal likelihood function is given by

$$\begin{aligned} L_{ML}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \{(2\pi)^{-\frac{mT_i}{2}} |\boldsymbol{V}_i(\boldsymbol{\alpha})|^{-\frac{1}{2}} \\ \times \exp(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}))\}. \end{aligned} \tag{4}$$

The marginal log-likelihood function $\ell(\boldsymbol{\theta})$ is then given by

$$\begin{aligned} \log L_{ML}(\boldsymbol{\theta}) = -\frac{mT}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \log |\boldsymbol{V}_i(\boldsymbol{\alpha})| \\ - \frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}), \end{aligned} \tag{5}$$

where $T = \sum_{i=1}^{n} T_i$ is the total number of observed outcomes from all the studies. We consider two scenarios below.

Assume $\boldsymbol{\alpha}$ is Known

We first consider a special case where $\boldsymbol{\alpha}$, and hence, $\boldsymbol{V}_i(\boldsymbol{\alpha})$ is known. This simple case is not very common, but it has important computational advantage of having a closed-form solution. Since $\boldsymbol{\alpha}$ is known, we are left with only $\boldsymbol{\beta}$ to estimate. This implies we only need to optimize or find the minimum of the function (West et al., 2014, Verbeke and Molenberghs, 2000)

$$\psi(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}), \tag{6}$$

which is the negative of the last term in Eq. (5). Optimization of $\psi(\boldsymbol{\beta})$ can be achieved using the method of generalized least squares (GLS) to get an analytical optimal value of

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{n}\boldsymbol{X}_i{}'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{Y}_i, \tag{7}$$

which is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ (West et al. (2014)).

Assume $\boldsymbol{\alpha}$ is Not Known

We consider the ML estimation for the variance and covariance parameters $\boldsymbol{\alpha}$ as well as fixed effects $\boldsymbol{\beta}$, in the case where $\boldsymbol{\alpha}$ is not known. We use the profile log-likelihood function $\ell(\boldsymbol{\alpha})$ to estimate the variance and covariance parameters where we replace the $\boldsymbol{\beta}$ parameters using Eq. (7) above. Thus,

$$\ell(\boldsymbol{\alpha}) = -\frac{mT}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{V}_i(\boldsymbol{\alpha})| - \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{r}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{r}_i, \tag{8}$$

where

$$\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i\left(\left(\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{Y}_i\right). \tag{9}$$

Maximization of $\ell(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ is a nonlinear optimization problem involving inequality constraints on $\boldsymbol{\alpha}$ to ensure that both $\boldsymbol{D}$ and $\boldsymbol{S}_i$ are positive-definite. Since there is no closed-form solution for the optimal $\boldsymbol{\alpha}$, it is estimated through a computational iterative procedure until convergence is reached. After obtaining the ML estimates of $\boldsymbol{\alpha}$ (and hence $\boldsymbol{D}$ and $\boldsymbol{S}_i$), $\boldsymbol{\beta}$ can now be estimated using the closed-form GLS formula in Eq. (7), that is,

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\boldsymbol{X}_i{}'\widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i'\widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{Y}_i, \tag{10}$$

where $\widehat{\boldsymbol{V}}_i = \boldsymbol{V}_i(\widehat{\boldsymbol{\alpha}}) = \boldsymbol{Z}_i\widehat{\boldsymbol{D}}\boldsymbol{Z}_i{}' + \widehat{\boldsymbol{S}}_i$. The solution in Eq. (10) above is called the empirical best linear unbiased estimator (EBLUE) of $\widehat{\boldsymbol{\beta}}$ since $\boldsymbol{V}_i$ has been replaced by its estimate $\widehat{\boldsymbol{V}}_i$. The variance of $\widehat{\boldsymbol{\beta}}$ is calculated by

$$\text{var}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \widehat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i \right)^{-1}. \tag{11}$$

However, the ML estimates of $\boldsymbol{\alpha}$ here are biased because they do not take account of the loss of degrees of freedom that results from estimating the fixed-effect parameters in $\boldsymbol{\beta}$ (Verbeke and Molenberghs, 2000). In order to eliminate this bias, the restricted maximum likelihood (REML) estimation method can be used, as described below.

3.2 Restricted Maximum Likelihood (REML) Estimation

REML estimation yields unbiased estimates of $\boldsymbol{\alpha}$ by taking account of the loss of degrees of freedom as a result of estimating $\boldsymbol{\beta}$. The estimates of $\boldsymbol{\alpha}$ are found by optimizing the REML log-likelihood function (West et al. (2014))

$$\begin{aligned} L_{REML}(\boldsymbol{\alpha}) = &-\frac{m}{2} \times (T - p) \times \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \log |\boldsymbol{V}_i| \\ &- \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{r}_i' \boldsymbol{V}_i^{-1} \boldsymbol{r}_i - \frac{1}{2} \sum_{i=1}^{n} \log |\boldsymbol{X}_i \boldsymbol{V}_i^{-1} \boldsymbol{X}_i|, \end{aligned} \tag{12}$$

where $\boldsymbol{r}_i$ is given by Eq. (9) above and p denotes the number of fixed effects in $\boldsymbol{\beta}$. Once estimates of $\boldsymbol{\alpha}$ are obtained, and hence $\widehat{\boldsymbol{V}}_i$, $\boldsymbol{\beta}$ is then estimated using Eq. (10) above, and the corresponding $\text{var}(\widehat{\boldsymbol{\beta}})$ is calculated by Eq. (11). Since $\boldsymbol{\alpha}$ is estimated differently, the estimates of $\boldsymbol{\beta}$ and the corresponding $\text{var}(\widehat{\boldsymbol{\beta}})$ from ML and REML are different. REML is preferred because it provides unbiased estimates of $\boldsymbol{\alpha}$.

4 Modeling Covariance Structures

For brevity and without loss of generality with regard to the real-data setting, we assume $T = 4$ time points and $m = 2$ correlated effect sizes for each study. We also assume, for parsimonious reasons and without loss of generality, that $\boldsymbol{X}_i$ consists of only time indicators such that $\boldsymbol{X}_i = \boldsymbol{I}_8$ (an identity matrix of order 8), where intercept terms are ignored. Four models with different covariance structures for (3) are proposed.

4.1 Model 1—Independent $m = 2$ Effect Sizes, Independent Random Time Effects, and Independent Within-Study Effect Sizes

In this model, we assume that the $m = 2$ effect sizes are independent, $\text{cov}(Y_{i1t}, Y_{i2t}) = 0$, and $\text{cov}(Y_{i1t}, Y_{i2t'}) = 0\ \forall t \neq t'$. We also assume that the corresponding random effects are independent, $\text{cov}(\delta_{i1t}, \delta_{i2t}) = 0$, and $\text{cov}(\delta_{i1t}, \delta_{i2t'}) = 0\ \forall t \neq t'$. We also assume that for each of the $m = 2$ effect sizes, the random effects are independent ($\text{cov}(\delta_{ilt}, \delta_{ilt'}) = 0\ \forall t \neq t'$), and the longitudinal effect sizes are independent, $\text{cov}(Y_{ilt}, Y_{ilt'}) = 0\ \forall t \neq t'$.

This model is therefore equivalent to performing separate univariate random-effect meta-analyses for each of the $m = 2$ effect sizes and also separately at each time point. This model allows independent random intercept effects for each study i, and each effect size l, and at each time point t, δ_{ilt}, such that

$$Y_{ilt} = \beta_{lt} + \delta_{ilt} + e_{ilt}, l = 1, 2; t = 1, \cdots, 4, \tag{13}$$

where we assume that $\delta_{ilt} \sim N(0, \tau_{lt}^2)$ and $e_{ilt} \sim N(0, \sigma_{ilt}^2)$ are independent. Let $\boldsymbol{Z}_i = \boldsymbol{X}_i = \boldsymbol{I}_8$ so that (3) becomes

$$\boldsymbol{Y}_i = \boldsymbol{\beta} + \boldsymbol{\delta}_i + \boldsymbol{e}_i,$$

and

$$V(\boldsymbol{Y}_i) = \boldsymbol{D} + \boldsymbol{S}_i = \text{diag}(\tau_{11}^2 + \sigma_{i11}^2, \cdots, \tau_{14}^2 + \sigma_{i14}^2, \tau_{21}^2 + \sigma_{i21}^2, \cdots, \tau_{24}^2 + \sigma_{i24}^2).$$

However, for each of the $m = 2$ effect sizes, this model ignores within-study serial correlation between longitudinal effect sizes that exists because it is the same individuals who are measured repeatedly at these time points. It also ignores the correlation between $m = 2$ correlated effect sizes observed at each time point.

4.2 Model 2—Correlated $m = 2$ Effect Sizes, Independent Random Time Effects, and Independent Within-Study Effect Sizes

This model is equivalent to performing separate multivariate meta-analyses of the $m = 2$ effect sizes at each time point. Thus, it takes account of the correlation between the two effect sizes such that $\text{cov}(Y_{i1t}, Y_{i2t}) = \rho_{s_{1t,2t}} \sigma_{i1t} \sigma_{i2t}$, where $\rho_{s_{1t,2t}} = \text{corr}(Y_{i1t}, and Y_{i2t})$ is the within-study correlation between effect sizes 1 and 2 at time point t (assumed the same for all studies). Also, σ_{i1t} and σ_{i2t} are the within-study standard deviations of effect sizes 1 and 2, respectively, at time point t. However, $\text{cov}(Y_{i1t}, Y_{i2t'}) = 0\ \forall t \neq t'$. The random time effects are also

correlated the same: $\text{cov}(\delta_{i1t}, \delta_{i2t}) = \rho_{\tau_{1t,2t}}\tau_{1t}\tau_{2t}$, with $\rho_{\tau_{1t,2t}} = \text{corr}(\delta_{i1t}, \delta_{i2t})$ is the between-study correlation between random effects for effect sizes 1 and 2, at time point t, and τ_{1t} and τ_{2t} are the between-study standard deviations for effect sizes 1 and 2, respectively, at time point t. However, $\text{cov}(\delta_{i1t}, \delta_{i2t'}) = 0\ \forall t \neq t'$. As in the independence model 1 above, the random effects at different time points are independent, that is $(\text{cov}(\delta_{ilt}, \delta_{ilt'}) = 0\ \forall t \neq t')$, and the longitudinal effect sizes are independent, $\text{cov}(Y_{ilt}, Y_{ilt'}) = 0\ \forall t \neq t'$ for $l = 1, 2$. To illustrate this model, the effect sizes vector $\boldsymbol{Y}_i$ is re-arranged to group the $m = 2$ effect sizes at each time point together, such that $\boldsymbol{Y}_i$ becomes

$$\begin{pmatrix} \begin{bmatrix} Y_{i11} \\ Y_{i21} \end{bmatrix} \\ \begin{bmatrix} Y_{i12} \\ Y_{i22} \end{bmatrix} \\ \begin{bmatrix} Y_{i13} \\ Y_{i23} \end{bmatrix} \\ \begin{bmatrix} Y_{i14} \\ Y_{i24} \end{bmatrix} \end{pmatrix}. \tag{14}$$

Assuming only time indicator variables implies $\boldsymbol{X}_i = \boldsymbol{Z}_i$ to be $mT_i \times mT_i$ identity matrix. The remaining terms, $\boldsymbol{\beta}$, $\boldsymbol{\delta}_i$, and $\boldsymbol{e}_i$, in (3) are defined similar to $\boldsymbol{Y}_i$. The variance–covariance matrix is symmetric $V(\boldsymbol{Y}_i) = \boldsymbol{D} + \boldsymbol{S}_i$, with symmetric $\boldsymbol{D}$ given by

$$\begin{pmatrix} \begin{bmatrix} \tau_{11}^2 & \rho_{\tau_{11,21}}\tau_{11}\tau_{21} \\ & \tau_{21}^2 \end{bmatrix} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \tau_{12}^2 & \rho_{\tau_{12,22}}\tau_{12}\tau_{22} \\ & \tau_{22}^2 \end{bmatrix} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \begin{bmatrix} \tau_{12}^2 & \rho_{\tau_{13,23}}\tau_{13}\tau_{23} \\ & \tau_{23}^2 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \begin{bmatrix} \tau_{14}^2 & \rho_{\tau_{14,24}}\tau_{14}\tau_{24} \\ & \tau_{24}^2 \end{bmatrix} \end{pmatrix}, \tag{15}$$

and also symmetric $\boldsymbol{S}_i$ similar to $\boldsymbol{D}$ except that τ is now replaced by σ_i and ρ_τ by ρ_s. Therefore in $\boldsymbol{D}$, we have $\rho_{\tau_{1t,2t}}$ as the between-study correlation, and in $\boldsymbol{S}_i$, we have $\rho_{s_{1t,2t}}$ as the within-study correlation. Since these two correlations are usually unknown, they are estimated in the model.

This model is equivalent to performing separate multivariate meta-analyses at each time point. Although this model accounts for dependence between the $m = 2$ effect sizes, it assumes independence between the longitudinal effect sizes, which are usually auto-correlated. There is need to account for the serial correlation between longitudinal measurements, and a model is required that takes account of this dependence.

4.3 Model 3—Independent $m = 2$ Effect Sizes, Correlated Random Time Effects, and Correlated Within-Study Effect Sizes

Although this model assumes that the $m = 2$ effect sizes are independent as explained in the independent model above, it accounts for the dependence between the effect sizes within each of the $m = 2$ effect sizes. A correlation (ρ_{τ_l}) between any two adjacent random effects for the lth effect size is assumed, and a heteroscedastic autoregressive structure of order one, HAR(1), in which the correlation between random time effects decays as the lag between them increases, $\text{corr}(\delta_{ilt}, \delta_{ilt'}) = \rho_{\tau_l}^{|t-t'|}$, $\forall t \neq t'$, is used. The same dependence between longitudinal effect sizes within the same study, namely, $\text{corr}(Y_{ilt}, Y_{ilt'}) = \rho_{s_l}^{|t-t'|}$, $\forall t \neq t'$, where ρ_{s_l} is the correlation between any two adjacent effect sizes for the same lth effect size, is also assumed. Therefore, the variance–covariance matrix is now given by $V(\boldsymbol{Y}_i) = \boldsymbol{D} + \boldsymbol{S}_i$,

$$\begin{pmatrix} \boldsymbol{V}_{1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{2,2} \end{pmatrix},$$

where $\boldsymbol{V}_{1,1}$ is a symmetric matrix, where the upper half matrix is given by

$$\begin{bmatrix} \tau_{11}^2 + \sigma_{i11}^2 & \rho_{\tau_1}\tau_{11}\tau_{12} + \rho_{s_1}\sigma_{i11}\sigma_{i12} & \rho_{\tau_1}^2\tau_{11}\tau_{13} + \rho_{s_1}^2\sigma_{i11}\sigma_{i13} & \rho_{\tau_1}^3\tau_{11}\tau_{14} + \rho_{s_1}^3\sigma_{i11}\sigma_{i14} \\ & \tau_{12}^2 + \sigma_{i12}^2 & \rho_{\tau_1}\tau_{12}\tau_{13} + \rho_{s_1}\sigma_{i12}\sigma_{i13} & \rho_{\tau_1}^2\tau_{12}\tau_{14} + \rho_{s_1}^2\sigma_{i12}\sigma_{i14} \\ & & \tau_{13}^2 + \sigma_{i13}^2 & \rho_{\tau_1}\tau_{13}\tau_{14} + \rho_{s_1}\sigma_{i13}\sigma_{i14} \\ & & & \tau_{14}^2 + \sigma_{i14}^2 \end{bmatrix}$$

and $\boldsymbol{V}_{2,2}$ is also a symmetric matrix, where the upper half matrix is given by

$$\begin{bmatrix} \tau_{21}^2 + \sigma_{i21}^2 & \rho_{\tau_2}\tau_{21}\tau_{22} + \rho_{s_2}\sigma_{i21}\sigma_{i22} & \rho_{\tau_2}^2\tau_{21}\tau_{23} + \rho_{s_2}^2\sigma_{i21}\sigma_{i23} & \rho_{\tau_2}^3\tau_{21}\tau_{24} + \rho_{s_2}^3\sigma_{i21}\sigma_{i24} \\ & \tau_{22}^2 + \sigma_{i22}^2 & \rho_{\tau_2}\tau_{22}\tau_{23} + \rho_{s_2}\sigma_{i22}\sigma_{i23} & \rho_{\tau_2}^2\tau_{22}\tau_{24} + \rho_{s_2}^2\sigma_{i22}\sigma_{i24} \\ & & \tau_{23}^2 + \sigma_{i23}^2 & \rho_{\tau_2}\tau_{23}\tau_{24} + \rho_{s_2}\sigma_{i23}\sigma_{i24} \\ & & & \tau_{24}^2 + \sigma_{i24}^2 \end{bmatrix}.$$

This model is equivalent to performing longitudinal meta-analysis, jointly for the $T = 4$ time points, but marginally for the $m = 2$ effect sizes. Although this model accounts for dependence between longitudinal effect sizes, it assumes independence between the $m = 2$ effect sizes measured at the same time point, which are usually correlated. For instance, in a meta-analysis of studies where the $m = 2$ outcomes of interest are viral load suppression and CD4 count, there is need to account for the correlation between these two measurements. A model is required that takes account of this dependence as well.

4.4 Model 4—Correlated $m = 2$ Effect Sizes at Each Time Point, Correlated Random Time Effects, and Correlated Within-Study Effect Sizes

This is an extension of the above model that allows both the within-study and between-study covariance of the $m = 2$ effect sizes measured at the same time point, however assuming zero covariance at different time points. Specifically, $\text{cov}(Y_{i1t}, Y_{i2t}) = \sigma_{i1t,i2t}$ and $\text{cov}(Y_{i1t}, Y_{i2t'}) = 0$, $\forall t \neq t'$. For the random time effects, $\text{cov}(\delta_{i1t}, \delta_{i2t}) = \tau_{1t,2t}$, and $\text{cov}(\delta_{i1t}, \delta_{i2t'}) = 0$, $\forall t \neq t'$. Like the model above, the dependence between longitudinal effect sizes is accounted for by a heteroscedastic AR(1) covariance structure both within and between studies, separately for each of the $m = 2$ effect sizes. The variance–covariance matrix is now given by $V(\boldsymbol{Y}_i) = \boldsymbol{D} + \boldsymbol{S}_i$,

$$\begin{pmatrix} \boldsymbol{V}_{1,1} \boldsymbol{V}_{1,2} \\ \boldsymbol{V}_{2,1} \boldsymbol{V}_{2,2} \end{pmatrix},$$

where $\boldsymbol{V}_{1,1}$ and $\boldsymbol{V}_{2,2}$ are the same as in the model above but $\boldsymbol{V}_{1,2} = \boldsymbol{V}_{2,1} = \text{diag}(\tau_{11,21} + \sigma_{i11,i21}, \tau_{12,22} + \sigma_{i12,i22}, \tau_{13,23} + \sigma_{i13,i23}, \tau_{14,24} + \sigma_{i14,i24})$.

5 Example: Antiretroviral Drugs in Treatment-Experienced HIV-Infected Patients

Pichenot et al. (2011)'s paper is a systematic review and meta-analysis of the efficacy of some antiretroviral drugs in treatment-experienced HIV-infected patients. There were 10 trials included in the meta-analysis with a total of 6401 HIV-infected patients. There were two univariate meta-analyses reported in Pichenot et al. (2011), namely the odds ratio of achieving viral load below 50 copies per ml and the mean difference in CD4 count change. Both effect sizes were measured at week 48, and the comparisons are between the intervention and control groups. In order to obtain further data for this chapter, full text articles for each of the included studies were retrieved. Complete data were successfully retrieved for 5 trials reporting these two outcomes at longitudinal time points (week 4, 8, 12, 16, 20, 24, 32, 40, and 48), and the data set is given in Table 1. Where standard errors (SE) were not given, we either calculated them from confidence intervals, estimated them from the graphs in the papers, or imputed them from other time points. We note that the imputations described in this paragraph may affect the meta-analysis results; however, the aim of this chapter is to compare models and not necessarily to show effectiveness of the interventions. A description of the data from each of the five studies is given below.

Suleiman et al. (2010) assessed vicriviroc (VCV) in combination therapy with an optimized regimen in the VICTOR-E1 intervention. There were two active treatment arms (VCV 20mg and VCV 30mg) versus the Placebo arm. The actual values for the

Table 1 Log OR (SE) for VL < 50 and MD (SE) for CD4 change in 5 trials

Study ID	Intervention	Week	Log OR (SE^2)	MD (SE^2)
Saag et al., 2009	A4001029	4	0.067 (0.403)	24.798 (273.599)
Saag et al., 2009	A4001029	8	0.006 (0.201)	11.798 (273.599)
Saag et al., 2009	A4001029	12	−0.168 (0.179)	25.817 (273.599)
Saag et al., 2009	A4001029	16	0.132 (0.157)	16.092 (273.599)
Saag et al., 2009	A4001029	20	−0.149 (0.166)	21.477 (273.599)
Saag et al., 2009	A4001029	24	1.897 (0.169)	24.477 (273.599)
Saag et al., 2009	A4001029	32		
Saag et al., 2009	A4001029	40		
Saag et al., 2009	A4001029	48		27.800 (273.599)
Gulick et al., 2008	MOTIVATE 1 & 2	4	0.496 (0.076)	32.029 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	8	0.635 (0.044)	43.043 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	12	0.789 (0.035)	46.000 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	16	0.338 (0.026)	50.029 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	20	1.008 (0.032)	53.014 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	24	1.008 (0.032)	50.957 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	32	1.288 (0.035)	54.986 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	40	1.211 (0.036)	69.014 (178.686)
Gulick et al., 2008	MOTIVATE 1 & 2	48	1.366 (0.039)	59.057 (178.686)
Suleiman et al., 2010	VICTOR-E1	4	0.677 (0.361)	26.500 (1148.034)
Suleiman et al., 2010	VICTOR-E1	8	0.746 (0.261)	53.500 (1148.034)
Suleiman et al., 2010	VICTOR-E1	12	0.988 (0.213)	52.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	16	1.036 (0.200)	53.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	20	1.060 (0.184)	50.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	24	1.498 (0.203)	42.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	32	1.043 (0.191)	66.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	40	1.188 (0.200)	74.000 (1148.034)
Suleiman et al., 2010	VICTOR-E1	48	1.969 (0.284)	56.000 (1148.034)
Hicks et al., 2006	RESIST - 1 & 2	4	0.334 (0.064)	20.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	8	0.649 (0.032)	25.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	12		
Hicks et al., 2006	RESIST - 1 & 2	16	0.910 (0.023)	28.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	20		
Hicks et al., 2006	RESIST - 1 & 2	24	0.850 (0.021)	29.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	32	0.867 (0.021)	28.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	40	0.928 (0.022)	28.000 (24.847)
Hicks et al., 2006	RESIST - 1 & 2	48	0.957 (0.022)	24.000 (24.847)
Clotet et al., 2007	POWER - 1 & 2	4	0.309 (0.247)	40.000 (89.78)
Clotet et al., 2007	POWER - 1 & 2	8	0.539 (0.146)	56.000 (89.78)
Clotet et al., 2007	POWER - 1 & 2	12	1.145 (0.113)	67.000 (98.18)
Clotet et al., 2007	POWER - 1 & 2	16	1.545 (0.108)	77.000 (131.38)
Clotet et al., 2007	POWER - 1 & 2	20	1.711 (0.108)	73.000 (131.38)
Clotet et al., 2007	POWER - 1 & 2	24	1.792 (0.107)	75.000 (150.49)
Clotet et al., 2007	POWER - 1 & 2	32	1.792 (0.107)	90.000 (139.78)
Clotet et al., 2007	POWER - 1 & 2	40	2.037 (0.120)	94.000 (191.69)
Clotet et al., 2007	POWER - 1 & 2	48	1.997 (0.120)	83.000 (207.85)

Log OR = logarithm odds ratio, SE = standard error, VL = viral load, MD = mean difference

two outcomes were given on the graphs given in Suleiman et al. (2010). For the first outcome, both numerator and denominator showing the proportion of responders with viral load below 50 copies per ml were given for the three treatment groups at weeks 0, 2, 4, 6, 8, 12, 16, 20, 24, 32, 40, and 48. Values for weeks 0, 2, and 6 were not used in the analysis since other studies did not report at these time points. Values for the two active treatment arms were combined, and log odds ratios were calculated with their corresponding standard errors (SE), with Placebo as the comparison group. The formula for standard error for the log odds ratio used was $\sqrt{1/a + 1/b + 1/c + 1/d}$, (Agresti, 1996) where the numbers a, b, c, d are entries of the corresponding 2×2 table for grouped outcome frequencies. An intention to treat analysis was assumed for this and all the included studies. For the second outcome, the values for the mean change in CD4 count from baseline were given on the graph for weeks 4, 8, 12, 16, 20, 24, 32, 40, and 48. Values for the two active treatment groups were combined into one. Since the standard errors for the mean difference in CD4 change were not given, values for week 48 derived from Pichenot et al. (2011) were assumed across all of these weeks. The mean difference for week 48 in Suleiman et al. (2010) of 56.0 was used in place of 45.5 from Pichenot et al. (2011).

Four included studies (Clotet et al., 2007 [POWER-1 & 2]; Gulick et al., 2008 [MOTIVATE-1 & 2], Hicks et al., 2006 [RESIST-1 & 2]; and Saag et al., 2009 [A4001029]) did not report the actual values but presented the graphics from which data were then extracted. For Clotet et al. (2007) [POWER-1 & 2] assessing darunavir–ritonavir (DRV/r) versus control (CPI(s)), the proportion of patients with viral load below 50 copies per ml was given in a graph for each week, and these proportions were extracted and used in calculating log odds ratios and corresponding standard errors. For the second outcome, values were extracted from a graph for mean change in CD4 count from baseline (cells per μL) together with standard errors for the mean changes that were presented as vertical bars at each week and for each treatment arm. The standard errors for the mean difference were then computed from the standard errors for the intervention and control groups. For Gulick et al. (2008) [MOTIVATE-1 & 2] assessing maraviroc (MVC) once daily and maraviroc (MVC) twice daily versus placebo, values were extracted from the graphs and results combined for the two active treatment arms. For the first outcome, the percentage of patients achieving viral load below 50 copies per ml was extracted for each week and treatment arm and were used to compute the log odds ratios and corresponding standard errors. For the second outcome, the mean differences were calculated from the values on mean change in CD4 count from baseline extracted from the graphs; however, the graph did not have vertical bars for standard errors, and therefore, the standard error for week 48 (from Pichenot et al. (2011)) was assumed for all the earlier weeks. For Hicks et al. (2006) [RESIST-1 & 2], assessing tipranavir–ritonavir (TPV/r) versus the control group (CPI/r), proportions were extracted from the graphs and used to calculate log odds ratios for the first outcome. For the second outcome, the mean CD4 change was extracted from the graphs, and since the standard errors were not given, the value for week 48 was used for the earlier weeks: this value was calculated from the standard deviations

given in the results section of Hicks et al. (2006). Finally for Saag et al., 2009 [A4001029], also assessing the maraviroc once daily and maraviroc twice daily versus placebo, values were extracted from the graphs like above and combined for the active treatment arms. Also, the standard error for the second outcome was taken from week 48 (from Pichenot et al. (2011)) and used for all the other earlier weeks: this study only assessed the two outcomes up to week 24, and values for week 48 for the second outcome were taken from Pichenot et al. (2011).

This example data set is used to illustrate the performance of the four models described above, and the results for the effect sizes and their corresponding 95% confidence intervals are summarized in Table 2. Values for the between-study variances and correlation estimates are given in the text. Restricted maximum likelihood (REML) estimation was used to estimate parameters via the SAS PROC MIXED procedure (see details in the SAS code in the Appendix). The SAS code was an adaptation or extension of the SAS code given in Ishak et al. (2007).

Results for Model 1

This model assumes independence between the two effect sizes and also between the random time effects, both within and between studies. It is the same as performing univariate random-effect meta-analyses at each time point and separately for each effect size. The results (Figs. 1, 2, and Table 2) show that the antiretroviral drugs significantly improved viral load suppression and increased the mean CD4 change compared to the control from week 4 to 48. For the first outcome, the log odds ratio increased from week 4 (log OR 0.395, 95%CI: 0.077, 0.713) to week 48 (log OR 1.456, 95%CI: 0.954, 1.958). Similarly, for the second outcome, the mean CD4 change increased from week 4 (MD 27.365, 95%CI: 16.097, 38.632) to week 40 (MD 63.254, 95% CI: 30.749, 95.759). The between-study variance estimates (heterogeneity) for the log odds ratio estimates, which are not shown in the table, ranged from 0 (week 4) to 0.428 (week 32); and from 44.499 (week 4) to 830.25 (week 40) for the second outcome. There are no correlation estimates for this model.

Results for Model 2

This model takes account of the correlation between the two effect sizes, log odds ratio of achieving viral load suppression, and the mean change in CD4 count, by utilizing separate multivariate meta-analyses at each time point. This correlation between viral load suppression and CD4 count is known to exist. Although the model was run at one instance, it is equivalent to performing separate multivariate random-effect meta-analyses at each week. This model, however, ignores the serial correlation between random time effects that exist both within and between studies.

As in the independence model 1 above, the odds ratios for achieving viral load suppression increased with time from week 4 (log OR 0.448, 95%CI: 0.157, 0.739) to week 48 (log OR 1.341, 95%CI: 0.934, 1.747). Comparing the log odds ratio estimates with the independence model 1, it is clear that the actual values are similar although the values from model 2 are generally larger and the 95% confidence intervals are more precise, probably due to the adjustment of this latter model 2 for the correlation between viral load and CD4 count. Similarly, this latter model

Table 2 Effect size estimates (95% CI) for each week and outcomes 1 and 2 using models 1 to 4 of meta-analysis of assessing antiretroviral drugs in 5 trials

Week		Model 1	Model 2	Model 3	Model 4
4	1	0.395 (0.077, 0.713)	0.448 (0.157, 0.739)	0.377 (0.202, 0.551)	0.390 (0.237, 0.542)
	2	27.365 (16.097, 38.632)	30.245 (18.129, 42.361)	24.433 (15.791, 33.074)	27.022 (21.263, 32.780)
8	1	0.593 (0.359, 0.828)	0.611 (0.384,, 0.839)	0.515 (0.295, 0.735)	0.591 (0.392, 0.790)
	2	36.053 (18.976, 53.129)	40.545 (23.963, 57.128)	30.784 (16.852, 44.717)	31.835 (20.881, 42.789)
12	1	0.723 (0.230, 1.216)	0.849 (0.292, 1.406)	0.731 (0.290, 1.172)	0.671 (0.238, 1.103)
	2	49.730 (29.172, 70.287)	58.219 (41.711, 74.727)	39.430 (27.092, 51.767)	40.299 (29.886, 50.713)
16	1	0.776 (0.302, 1.251)	0.696 (0.504, 0.888)	0.792 (0.262,1.322)	0.699 (0.239, 1.159)
	2	43.980 (20.867, 67.092)	43.936 (23.571, 64.301)	36.039 (20.516, 51.562)	37.232 (22.701, 51.762)
20	1	0.933 (0.210, 1.657)	0.942 (0.208, 1.677)	0.825 (-0.144, 1.794)	0.880 (0.289, 1.470)
	2	51.310 (26.867, 75.754)	51.964 (30.585, 73.343)	38.196 (25.0, 51.357)	38.970 (26.163, 51.778)
24	1	1.308 (0.876, 1.740)	1.299 (0.935, 1.663)	1.409 (0.821, 1.997)	1.084 (0.733, 1.434)
	2	44.252 (23.671, 64.833)	48.788 (31.610, 65.966)	36.694 (23.857, 49.532)	37.713 (24.686, 50.741)
32	1	1.207 (0.815, 1.599)	1.245 (0.819, 1.671)	1.305 (0.926, 1.685)	1.027 (0.422, 1.634)
	2	57.487 (27.044, 87.929)	55.258 (25.870, 84.645)	38.934 (21.347, 56.521)	39.701 (20.975, 58.428)
40	1	1.283 (0.832, 1.735)	1.206 (0.833, 1.530)	1.392 (0.952, 1.831)	1.056 (0.366, 1.745)
	2	63.254 (30.749, 95.759)	59.344 (30.711, 87.976)	40.280 (22.190, 58.370)	42.046 (23.324, 60.769)
48	1	1.456 (0.954, 1.958)	1.341 (0.934, 1.747)	1.616 (1.123, 2.109)	1.150 (0.438, 1.863)
	2	47.911 (23.383, 72.439)	48.252 (24.163, 72.341)	34.852 (19.240, 49.932)	35.809 (19.848, 51.771)
AIC		387.6	362.0	297.4	289.0

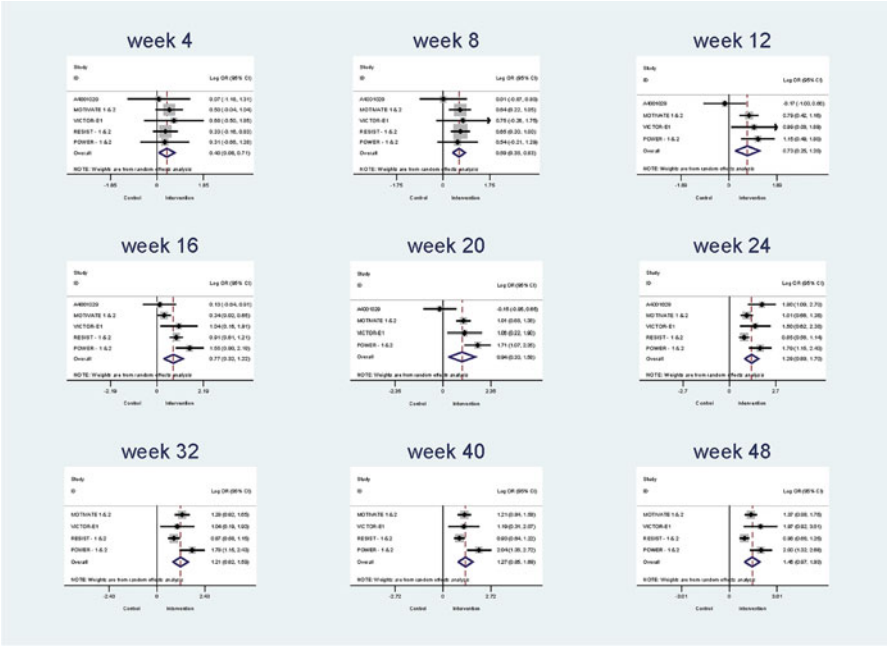


Fig. 1 Forest plot showing meta-analysis of log odds ratios for achieving viral load suppression for 5 trials assessing antiretroviral drugs in HIV treatment-experienced patients

2 yielded similar but slightly larger mean difference in CD4 count that had more precise confidence intervals; the values ranged from week 4 (MD 30.245, 95%CI: 18.129, 42.361) to week 40 (MD 59.344, 95% CI: 30.711, 87.976). Values for the between-study variances ranged from 0 (week 4 and 16) to 0.491 (week 20) for the first outcome and from 74.236 (week 4) to 688.54 (week 32) for the second outcome. All the nine between-study correlations were significantly high with a value of 1, except for week 4 that had a value of −1. The within-study correlations were ones for weeks 8, 12, and 20, ranged from −1, 0.635, 0.732, 0.820, 0.842, and 0.917 for weeks 24, 16, 48, 40, 32, and 4, respectively.

Model 2 performed better than model 1 as shown by tighter confidence intervals and lower AIC value (362.0 versus 387.6 for model 1). This benefit is attributed to accounting for the correlation between the two outcomes.

Results for Model 3

This model takes account of the serial correlation between the longitudinal effect sizes, both through the actual effect sizes and via the random time effects. The serial correlation exists because the measurements are taken repeatedly from the same study population. However, this model does not take account of the correlation between the two outcomes that were adjusted for in model 2 above. Therefore, this model accounts for the serial auto-correlations separately for the two outcomes. The

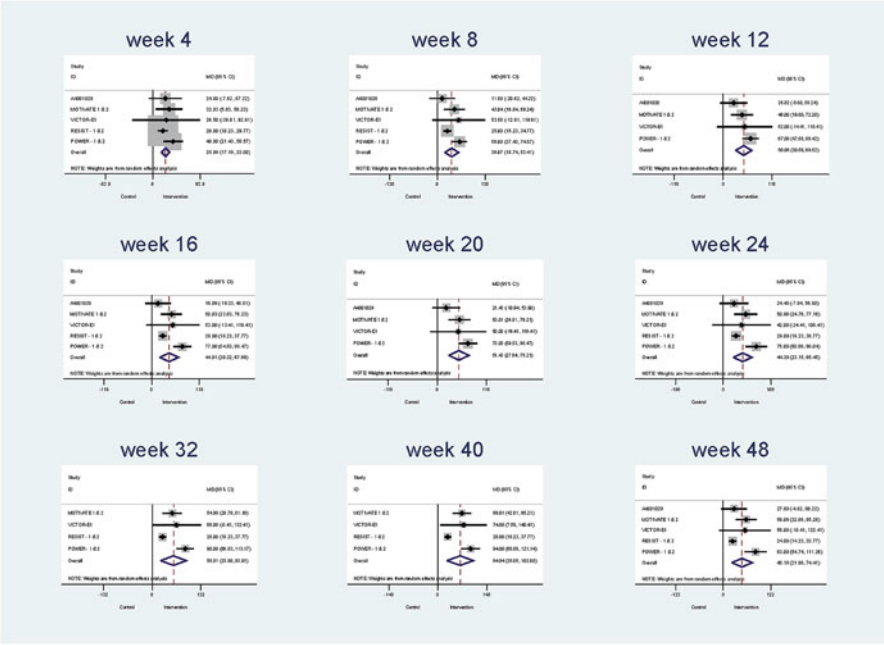


Fig. 2 Forest plot showing meta-analysis of mean difference in CD4 change for 5 trials assessing antiretroviral drugs in HIV treatment-experienced patients

parameter estimates for this model are generally similar to the independence model 1 above except that the 95% confidence intervals are generally more precise due to the adjustment for the serial correlation. The variance estimates were generally smaller compared to the independence model. The between-study correlations obtained were 0.773 (95% CI: 0.553, 0.993) for the first outcome (ρ_{τ_1}) and 0.979 (95% CI: 0.951, 1.000) for the second outcome (ρ_{τ_2}); however, the within-study correlations (ρ_{s_1} and ρ_{s_2}) were almost zero.

The benefit of accounting for the serial auto-correlation between longitudinal effect sizes, both within and between studies, resulted in the gain in precision of the parameter estimates. Using AIC, where smaller values are better, it is clear that model 3 (AIC 297.4) performed much better compared to model 2 (AIC 362.0) and model 1 (AIC 387.6).

Results for Model 4

This model is an extension of the model 3 above, which allows a non-zero covariance between the $m = 2$ outcomes at each time point, both between and within studies. Inspection of the model 4 results in Table 2 shows that although the parameter estimates were similar to model 3, they were all smaller than the independence model, and the confidence intervals were much more precise compared to all the other models 1 to 3. The same pattern of results is evident

here with parameter estimates for the first outcome increasing from week 4 to week 48, while values for the second outcome increased from week 4 to week 40. The between-study variance estimates and correlations were similar to model 3. The additional covariance estimates for this model were ranging from 0 to 16.4 although the majority of the estimates were zero due to the over-parametrization of the model.

The AIC value of 289.0 is the smallest compared to the rest of the models. This, coupled with the tighter confidence intervals, implies that model 4 performed best compared to the rest of the models, at least for this data set.

6 Summary

In this chapter, the longitudinal meta-analysis of multiple effect sizes was described, where four models with different covariance structures for random time effects and residuals, both within and between studies, to account for the correlation between effect sizes, were proposed. The models were compared using a practical example involving longitudinal meta-analysis of HIV studies reporting the log odds ratio of achieving viral load suppression and the mean CD4 change at 4, 8, 12, 16, 20, 24, 32, 40, and 48 weeks after start of treatment. The model that accounted for both correlations, namely between the two outcomes and the serial correlations between the longitudinal effect sizes, performed best among the four models. Although the values of the parameter estimates were similar between the four models, precision in terms of the width of the 95% confidence intervals improved from model 2 to model 4 compared to the independence model.

Extensions for this chapter are possible. The models proposed can be extended to include other covariance structures. These models also need to be validated through simulation studies. In addition, since the modeling approach used here involved estimating point estimates at each fixed time point, it is possible to also consider treating time as a continuous covariate and explore both linear and nonlinear models as shown in Ahn and French (2010). The other aspect that can be extended is to relax the normality assumption of the effect estimates and consider non-normal distributions. The Bayesian approach to estimating parameters for longitudinal meta-analysis is also another area of potential extension, as proposed in Lopes et al. (2003). Furthermore, the meta-analytic models proposed in this chapter use aggregate data from studies, and methods for patient-level data are needed as these can improve the power and precision of parameter estimates, as proposed in Farlow et al. (2005). The patient-level models can potentially allow the estimation of within-study correlations needed in the estimation of parameters. Finally, the bootstrap technique (Davison and Hinkley, 1997) can improve the precision of parameter estimates.

Acknowledgments This chapter is part of a PhD in Statistics entitled “Meta-analysis of longitudinal studies with multiple effect sizes” by one of the chapter authors (Alfred Musekiwa). The full PhD thesis can be accessed from the University of KwaZulu-Natal Library.

Appendix: SAS Code

```
*SAS Code
*Model 1
libname in "c:\analysis\paper4\new_paper4_analysis";
proc import
datafile="c:\analysis\paper4\new_paper4_analysis\data_sas_pichenot_final.xls"
dbms=xls out=data_final;
run;
data 'c:\analysis\paper4\new_paper4_analysis\data_final';
set data_final;
run;
data data_final;
set data_final;
w=1/var_effect;
run;
data within_study_cov;
param="Var_41"; est=1; output;
param="Var_81"; est=1; output;
param="Var_121"; est=1; output;
param="Var_161"; est=1; output;
param="Var_201"; est=1; output;
param="Var_241"; est=1; output;
param="Var_321"; est=1; output;
param="Var_401"; est=1; output;
param="Var_481"; est=1; output;
param="Var_42"; est=1; output;
param="Var_82"; est=1; output;
param="Var_122"; est=1; output;
param="Var_162"; est=1; output;
param="Var_202"; est=1; output;
param="Var_242"; est=1; output;
param="Var_322"; est=1; output;
param="Var_402"; est=1; output;
param="Var_482"; est=1; output;
param="Corr"; est=0; output;
run;
data btw_study_cov;
param="tau41"; est=0.100; output;
param="tau81"; est=0.100; output;
param="tau121"; est=0.1222; output;
param="tau161"; est=0.1784; output;
param="tau201"; est=0.3185; output;
param="tau241"; est=0.1266; output;
param="tau321"; est=0.0912; output;
param="tau401"; est=0.1115; output;
param="tau481"; est=0.1549; output;
param="tau42"; est=0.100; output;
param="tau82"; est=214.813; output;
param="tau122"; est=152.1187; output;
param="tau162"; est=501.4165; output;
param="tau202"; est=303.8539; output;
param="tau242"; est=363.9144; output;
param="tau322"; est=1000; output;
param="tau402"; est=1200; output;
param="tau482"; est=636.4071; output;
param="Corr"; est=0; output;
run;
data initial_values;
set btw_study_cov within_study_cov;
keep param est; run;
```

```
proc print data=initial_values;
run;
proc mixed method=REML cl
data=data_final;
class study_id week outcome;
model effect_size=week*outcome
/noint s cl ddf=1000,1000,
1000,1000, 1000,1000,
1000,1000, 1000, 1000,1000,
1000,1000, 1000,1000,
1000,1000, 1000;
random week (outcome)
/subject=study_id  type=arh(1);
repeated week (outcome)
/subject= study_id type=arh(1);
parms/parmsdata=initial_values
hold=19 to 38;
weight w;
run;

*Model 2
libname in "c:\analysis\paper4\new_paper4_analysis";
proc import
datafile="c:\analysis\paper4\new_paper4_analysis\data_sas_pichenot_final.xls"
dbms=xls out=data_final;
run;
data 'c:\analysis\paper4\new_paper4_analysis\data_final';
set data_final;
run;
data data_final;
set data_final;
w=1/var_effect;
run;
data btw_study_cov;
param="Var_21"; est=2; output;
param="Var_22"; est=2; output;
param="Corr21"; est=0.5; output;

param="Var_21"; est=2; output;
param="Var_22"; est=200; output;
param="Corr21"; est=0.5; output;

param="Var_21"; est=2; output;
param="Var_22"; est=200; output;
param="Corr21"; est=0.9; output;

param="Var_21"; est=100; output;
param="Var_22"; est=500; output;
param="Corr21"; est=0.9; output;

param="Var_21"; est=2; output;
param="Var_22"; est=300; output;
param="Corr21"; est=0.5; output;

param="Var_21"; est=2; output;
param="Var_22"; est=300; output;
param="Corr21"; est=0.5; output;

param="Var_21"; est=2; output;
```

```
param="Var_22"; est=1000; output;
param="Corr21"; est=0.5; output;

param="Var_21"; est=2; output;
param="Var_22"; est=1200; output;
param="Corr21"; est=0.1; output;

param="Var_21"; est=2; output;
param="Var_22"; est=700; output;
param="Corr21"; est=0.5; output;
run;
data within_study_cov;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.5; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.5; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.5; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.1; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.1; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.5; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.5; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.1; output;
param="Var_21"; est=1; output;
param="Var_22"; est=1; output;
param="Corr21"; est=0.1; output;
run;
data initial_values;
set btw_study_cov within_study_cov;
keep param est;
run;
proc mixed method=REML cl
data=data_final;
class study_id week outcome;
model effect_size=week*outcome
/noint s cl ddf=1000,1000;
random outcome
/subject=study_id group=week type=arh(1);
repeated outcome
/subject=study_id group=week type=arh(1);
parms/parmsdata=initial_values
hold=28 to 29, 31 to 32, 34 to 35, 37 to 38, 40 to 41, 43 to 44, 46 to 47, 49 to
50, 52 to 53;
weight w;
run;
```

```
*Model 3

libname in "c:\analysis\paper4\new_paper4_analysis";
proc import
datafile="c:\analysis\paper4\new_paper4_analysis\data_sas_pichenot_final.xls"
dbms=xls out=data_final;
run;
data 'c:\analysis\paper4\new_paper4_analysis\data_final';
set data_final;
run;
data data_final;
set data_final;
w=1/var_effect;
run;
data btw_study_cov;
param="tau41"; est=2; output;
param="tau81"; est=2; output;
param="tau121"; est=2; output;
param="tau161"; est=2; output;
param="tau201"; est=2; output;
param="tau241"; est=2; output;
param="tau321"; est=2; output;
param="tau401"; est=2; output;
param="tau481"; est=2; output;
param="Corr"; est=0.9; output
param="var_41"; est=0; output;
param="var_81"; est=0; output;
param="var_121"; est=0; output;
param="var_161"; est=0; output;
param="var_201"; est=0; output;
param="var_241"; est=0; output;
param="var_321"; est=0; output;
param="var_401"; est=0; output;
param="var_481"; est=0; output;
param="corr"; est=0; output;
param="var_41"; est=0; output;
param="var_81"; est=0; output;
param="var_121"; est=0; output;
param="var_161"; est=0; output;
param="var_201"; est=0; output;
param="var_241"; est=0; output;
param="var_321"; est=0; output;
param="var_401"; est=0; output;
param="var_481"; est=0; output;
param="corr"; est=0; output;
param="tau42"; est=0.100; output;
param="tau82"; est=214.813; output;
param="tau122"; est=152.1187; output;
param="tau162"; est=501.4165; output;
param="tau202"; est=303.8539; output;
param="tau242"; est=363.9144; output;
param="tau322"; est=1000; output;
param="tau402"; est=1200; output;
param="tau482"; est=636.4071; output;
param="Corr"; est=0.9; output;
run;
data within_study_cov;
param="Var_41"; est=1; output;
param="Var_81"; est=1; output;
```

```
param="Var_121"; est=1; output;
param="Var_161"; est=1; output;
param="Var_201"; est=1; output;
param="Var_241"; est=1; output;
param="Var_321"; est=1; output;
param="Var_401"; est=1; output;
param="Var_481"; est=1; output;
param="Corr"; est=0.9; output;
param="var_41"; est=0; output;
param="var_81"; est=0; output;
param="var_121"; est=0; output;
param="var_161"; est=0; output;
param="var_201"; est=0; output;
param="var_241"; est=0; output;
param="var_321"; est=0; output;
param="var_401"; est=0; output;
param="var_481"; est=0; output;
param="corr"; est=0; output;
param="var_41"; est=0; output;
param="var_81"; est=0; output;
param="var_121"; est=0; output;
param="var_161"; est=0; output;
param="var_201"; est=0; output;
param="var_241"; est=0; output;
param="var_321"; est=0; output;
param="var_401"; est=0; output;
param="var_481"; est=0; output;
param="corr"; est=0; output;
param="Var_42"; est=1; output;
param="Var_82"; est=1; output;
param="Var_122"; est=1; output;
param="Var_162"; est=1; output;
param="Var_202"; est=1; output;
param="Var_242"; est=1; output;
param="Var_322"; est=1; output;
param="Var_402"; est=1; output;
param="Var_482"; est=1; output;
param="Corr"; est=0.9; output;
run;
data initial_values;
set btw_study_cov within_study_cov;
keep param est;
run;
proc mixed method=REML cl
data=data_final;
class study_id week outcome;
model effect_size=outcome*week
/noint s cl ddf=1000,1000,
1000,1000, 1000,1000,
1000,1000, 1000;
random week
/subject=study_id group=outcome type=arh(1);
repeated week
/subject=study_id group=outcome type=arh(1);
parms/parmsdata=initial_values
hold=11 to 30, 41 to 49, 51 to 79;
weight w;
run;
```

```
*Model 4

libname in "c:\analysis\paper4\new_paper4_analysis";
proc import
datafile="c:\analysis\paper4\new_paper4_analysis\data_sas_pichenot_final.xls"
dbms=xls out=data_final;
run;
data 'c:\analysis\paper4\new_paper4_analysis\data_final';
set data_final;
run;
data data_final;
set data_final;
w=1/var_effect;
run;
data btw_study_cov;
param="Var_21"; est=20; output;
param="Var_23"; est=20; output;
param="Var_21"; est=20; output;
param="Var_22"; est=20; output;
param="Var_23"; est=20; output;
param="Var_24"; est=20; output;
param="Var_24"; est=20; output;
param="Var_24"; est=20; output;
param="Var_24"; est=20; output;
param="Corr21"; est=0.9; output;
param="Var_21"; est=50; output;
param="Var_23"; est=80; output;
param="Var_23"; est=160; output;
param="Var_23"; est=180; output;
param="Var_21"; est=200; output;
param="Var_22"; est=220; output;
param="Var_23"; est=420; output;
param="Var_24"; est=580; output;
param="Var_24"; est=500; output;
param="Corr21"; est=0; output;
param="Var_21"; est=50; output;
param="Var_23"; est=80; output;
param="Var_23"; est=160; output;
param="Var_23"; est=180; output;
param="Var_21"; est=200; output;
param="Var_22"; est=220; output;
param="Var_23"; est=420; output;
param="Var_24"; est=580; output;
param="Var_24"; est=500; output;
param="Corr21"; est=0; output;
param="tau42"; est=10; output;
param="tau42"; est=214.813; output;
param="tau122"; est=152.1187; output;
param="tau122"; est=501.4165; output;
param="tau202"; est=303.8539; output;
param="tau242"; est=363.9144; output;
param="tau322"; est=1000; output;
param="tau402"; est=1200; output;
param="tau482"; est=636.4071; output;
param="Corr"; est=0.9; output;
run;
data within_study_cov;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
```

```
param="Var_11"; est=1; output;
param="Var_12"; est=1; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Corr11"; est=0.2; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_11"; est=1; output;
param="Var_12"; est=1; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Corr11"; est=0; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_11"; est=1; output;
param="Var_12"; est=1; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Corr11"; est=0; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_11"; est=1; output;
param="Var_12"; est=1; output;
param="Var_13"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Var_14"; est=1; output;
param="Corr11"; est=0.9; output;
run;
data initial_values;
set btw_study_cov within_study_cov;
keep param est;
run;
proc mixed method=REML cl
data=data_final;
class study_id week outcome;
model effect_size=outcome*week
/noint s cl ddf=1000,1000,
1000,1000, 1000,1000,
1000, 1000, 1000;
random week
/subject=study_id group=outcome type=arh(1);
repeated week
/subject=study_id group=outcome type=arh(1);
parms/parmsdata=initial_values
hold=20,30, 41 to 49, 51 to 79;
weight w;
run;
```

References

Agresti, A. (1996). An introduction to categorical data analysis. New York: John Wiley & Sons, Inc.

Ahn, J. E., & French, J. L. (2010). Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *Journal of Pharmacokinetics and Pharmacodynamics, 37*(2), 179–201. https://doi.org/10.1007/s10928-010-9152-6. PMID: 20358280.

Arends, L. R., Vokó, Z., & Stijnen, T. (2003). Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine, 22*(8), 1335–1353.

Bax, L., Yu, L. M., Ikeda, N., Tsuruta, N., Moons, K. G. M. (2006). *MIX: Comprehensive Free Software for Meta-analysis of Causal Research Data—Version 1.5*.

Berkey, C. S., Anderson, J. J., & Hoaglin, D. C. (1996). Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine, 15*(5), 537–557.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine, 17*(22), 2537–2550.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, U.K.: Wiley.

Bradburn, M., Deeks, J., Altman, D. (1998). Metan—an alternative meta-analysis command. *Stata Technical Bulletin, 44*, 1–44.

Chevarier, P., Cucherat, M., Freiburger, T., Maupas, J., Visele, N., Bugnard, F., Bazog, P., & Weasy, M. A. (2000). *Lyon*.

Clotet, B., Bellos, N., Molina, J. M., Cooper, D., Goffard, J. C., Lazzarin, A., Wöhrmann, A., Katlama, C., Wilkin, T., Haubrich, R., & Cohen, C. (2007). Efficacy and safety of darunavir-ritonavir at week 48 in treatment-experienced patients with HIV-1 infection in POWER 1 and 2: a pooled subgroup analysis of data from two randomised trials. *The Lancet, 369*(9568), 1169–1178.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The Handbook of Research Synthesis and Meta-analysis* (2nd edn.). New York: Russell Sage Foundation.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Vol. 1). Cambridge: Cambridge University Press.

Egger, M., Smith, G. D., Altman, D. (Eds.). (2008). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd edn.). New York: Wiley.

Farlow, M. R., Small, G. W, Quarg, P, & Krause, A. (2005). Efficacy of Rivastigmine in Alzheimer's disease patients with rapid disease progression: results of a meta-analysis. *Dement Geriatr Cogn Disord, 20*, 192–197.

Fine, H. A., Dear, K. B., Loeffler, J. S., Black, P. M., & Canellos, G. P. (1993). Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults. *Cancer, 71*(8), 2585–2597.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper, & L. V. Hedges (Eds.). *The handbook of research synthesis and meta-analysis* (pp. 339-355). New York: Russell Sage Foundation.

Gulick, R. M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A., Wohlfeiler, M., & Montana, J. B. (2008). Maraviroc for previously treated patients with R5 HIV-1 infection. *New England Journal of Medicine, 359*(14), 1429–1441.

Hedges, L. V. & Olkin I. (1985). *Statistical Methods for Meta-analysis*. New York: Academic Press.

Hicks, C. B., Cahn, P., Cooper, D. A., Walmsley, S. L., Katlama, C., Clotet, B., Lazzarin, A., Johnson, M.A., Neubacher, D., Mayers, D., & Valdez, H. (2006). Durable efficacy of tipranavir-ritonavir in combination with an optimised background regimen of antiretroviral drugs for

treatment-experienced HIV-1-infected patients at 48 weeks in the Randomized Evaluation of Strategic Intervention in multi-drug reSistant patients with tipranavir (RESIST) studies: an analysis of combined data from two randomised open-label trials. *The Lancet, 368* (9534), 466–475.

Higgins, J. P. T., & Green, S. (Eds.). (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley.

Ishak, K. J., Platt, R. W., Joseph, L., Hanley, J. A., Caro, I. J. (2007). Meta-analysis of longitudinal studies. *Clinical Trials, 4*, 525–539.

Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine, 30*(20), 2481–2498.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*(3), 227–235.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963–974.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis*. Thousand Oaks, CA: Sage.

Litell, J., Corcoran, J., Pillai, V. (2008). *Systematic Reviews and Meta-analysis*. New York: Oxford University Press.

Lopes, H. F., Muller, P., Rosner, G. L. (2003). Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics, 59*, 66–75.

Mavridis, D., & Salanti, G. (2012). A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research, 22*(2), 1–26.

Musekiwa, A., Manda, S. O., Mwambi, H. G., & Chen, D. G. (2016). Meta-Analysis of Effect Sizes Reported at Multiple Time Points Using General Linear Mixed Model. *PLoS One*, e0164898. https://doi.org/10.1371/journal.pone.0164898.

Nam, I. S., Mengersen, K., & Garthwaite, P. (2003). Multivariate meta-analysis. *Statistics in Medicine, 22*(14), 2309–2333.

Pichenot, M., Deuffic-Burban, S., Cuzin, L., & Yazdanpanah, Y. (2011). Efficacy of new antiretroviral drugs in treatment-experienced HIV-infected patients: a systematic review and meta-analysis of recent randomized controlled trials. *HIV Medicine, 13*(3), 148–155.

Pigott, T. D. (2012). *Advances in Meta-analysis*. New York: Springer.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin, 103*(1), 111–119.

Review Manager (RevMan) (2014). [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.

Riley, R. D. (2009). Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society, Series A, 172*(4), 789–811.

Rosenberg, M. S., Adams, D. C., Gurevitch, J. (2000). *Statistical Software for Meta-Analysis Version 2*. Massachusetts: Sunderland.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*(3), 400–406.

Saag, M., Goodrich, J., Fätkenheuer, G., Clotet, B., Clumeck, N., Sullivan, J., Westby, M., Van Der Ryst, E., & A4001029 Study Group. (2009). A double-blind, placebo-controlled trial of maraviroc in treatment-experienced patients infected with non-R5 HIV-1. *Journal of Infectious Diseases, 199*(11), 1638–1647.

SAS Institute Inc. (2013). *SAS/STAT 9.3 user's guide*. Cary, NC: SAS Institute Inc.

Suleiman, J., Zingman, B. S., Diaz, R. S., Madruga, J. V. R., DeJesus, E., Slim, J., Mak, C., Lee, E., McCarthy, M. C., Dunkle, L. M., & Walmsley, S. (2010). Vicriviroc in combination therapy with an optimized regimen for treatment-experienced subjects: 48-week results of the VICTOR-E1 phase 2 trial. *Journal of Infectious Diseases, 201*(4), 590–599.

Sutton, A. J., Abrams, K. A., Jones, D. R., Sheldon, T. A., Song, F. (2000). *Methods for Meta-analysis in Medical Research*. Chichester, UK: Wiley.

Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine, 21*(4), 589–624.

Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Verbeke, G., Fieuws, S., Molenberghs, G., & Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research, 23*(1), 42–59.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software **36**(3), 1–48.

Wallace, B. C., Schimd, C. H., Lau, J.,Trikalinos, T. A. (2009). Meta-analyst: software for meta-analysis of binary, continuous and diagnostic data. BMC Medical Research Methodology **9**(1), 80. https://doi.org/10.1186/1471-2288-9-80.

Wallace, B. C., Dahabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., Schimd, C. H. (2012). Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. *Journal of Statistical Software, 49*(5), 1–15.

West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: A practical guide using statistical software* (2nd edn.). New York: Chapman and Hall/CRC.

Whitehead, A. (2002). *Meta-analysis of Controlled Clinical Trials*. New York: Wiley.

Part III
Spatial-Temporal Modelling and Disease Mapping

Measuring Bivariate Spatial Clustering in Disease Risks

Timotheus B. Darikwa and Samuel O. M. Manda

Abstract In most analyses of spatial variation in relative disease risk, consideration is on raised disease risk in each area. However, interest may also be in the wider clustering pattern across neighbouring areas, especially for health prioritisation. For a single disease, an area could be classified as high or low risk and may or may not be consistent with risk levels in the neighbouring areas. Recent developments in spatial clustering measures have been concerned with bivariate spatial autocorrelation measures. In this chapter, we present some spatial statistical approaches to measuring bivariate spatial clustering for bivariate spatial outcomes. Our application considers univariate spatial clustering in cerebrovascular (CVA) or ischaemic heart failure (IHD) or hypertension (HHD) and diabetes (DBT) deaths among the 30–70-year old in South Africa. The analysis is extended to bivariate spatial clustering when two causes of death are considered at a time.

Keywords Univariate spatial autocorrelation · Bivariate spatial autocorrelation · Mortality · Cardiovascular condition · South Africa

1 Introduction

In spatial statistics, spatial clustering statistical methods have long been used to group spatial objects into groups called clusters, so that objects in one cluster have similar characteristics compared to objects in other clusters. Most of the development in spatial clustering methods have focused on one areal health data (outcome), and the most widely used measure is the Moran's I index of spatial

T. B. Darikwa (✉)
Department of Statistics and Operations Research, University of Limpopo, Sovenga, South Africa
e-mail: Timotheus.Darikwa@ul.ac.za

S. O. M. Manda
Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: samuel.manda@up.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_9

autocorrelation (Moran, 1950). Both local and global indexes are widely available and implemented in many geographic information system (GIS) computer packages (Anselin, 1995; Anselin et al., 2002; Waller and Gotway, 2004). However, based on the original work of Mantel (1967), the univariate Moran's *I* has recently been expanded to cases where there are two spatially measured health data.

Wartenberg (1985) was the first to conceive the idea of using a bivariate spatial autocorrelation matrix to determine a multivariate spatial autocorrelation measure for more than two variables. Prior to that, spatial autocorrelation (SAC), which is a measure of how a variable correlates with itself over geographic space, was widely done for only one variable (univariate) as opposed to at least two variables (multivariate). The Moran's *I* and Geary's *C* have been by far the most popular univariate measures of spatial autocorrelation in various applications (White et al., 1989; Zhang et al., 2008; Matkan et al., 2013).

The method suggested by Wartenberg (1985) for extending Moran's I to multivariate spatial analysis involves the derivation of a matrix of bivariate SACs. This matrix is, in turn, analysed using spatial principal component analysis (sPCA) resulting in a set of spatial factors that represent the total spatial pattern. His formulation was criticised by Lee (2001) for its susceptibility to changes in the direction of the spatial association. In addition, (Lee, 2001) showed that the bivariate SAC matrix derived by Wartenberg (1985) may be asymmetric that is problematic when deriving the total multivariate spatial pattern but not impossible (Dray et al., 2008). In addition, (Lee, 2001) conceptualised that a bivariate SAC measure must be a function of the respective individual univariate spatial autocorrelation and the "point to point" correlation of some sort between the two variables as measured by Pearson's correlation coefficient. He further showed that the bivariate measure derived by Lee (2001) was a function of only one univariate SAC and a correlation between one variable and the lag of the second variable. Lee (2001) then derived a bivariate SAC measure, using a row-standardised weight matrix, that is in line with his conditions for a bivariate SAC measure and also produces a symmetric bivariate SAC matrix to be used for deriving the total multivariate spatial autocorrelations. The Pearson's correlation part of his derivation is between the spatial lags of the two variables that will be under consideration. One problem that may arise is that this correlation may differ significantly from that between the original values of the two variables and may have different signs (Lee, 2001; Dray et al., 2008). Lee (2001) also came up with univariate spatial measure known as Lee's *S* that can be used to calculate univariate spatial autocorrelation for one health outcome.

Bivariate measures need to be implemented to real-life problems that are currently lacking. There have been few applications of Lee's bivariate SAC with real data to date (Khamis, 2012; Khamis et al., 2014). Khamis (2012) investigated the bivariate spatial autocorrelation between unemployment rate (UR) and chronic illnesses (CI) in Iraq. They found no evidence of joint spatial autocorrelation between UR and CI, but individually the former tended to cluster but with no sufficient evidence for the clustering of the later. The bivariate relationship between mortality rate (MR) and socioeconomic factors as represented by UR, crime rate (CR), and divorce rate (DR) was investigated by Khamis et al. (2014) in Jordan

using Lee's formulation. Lee's global spatial autocorrelation was not found to be significant, but the local indicator was used to deduce, among other results, that mortality hotspots were spatially associated with hotspots of economically disadvantaged areas.

Recent availability of interrelated health outcomes data in South Africa such as cause of deaths data from Statistics South Africa means that there is a need to apply multivariate spatial association measures to identify co-clustering of health outcomes. This chapter uses 2001 and 2011 cause of deaths data to derive estimates of the univariate and bivariate measures for deaths due to cerebrovascular heart disease, ischaemic heart disease, hypertensive heart disease, and diabetes in South Africa based on the formulations of Lee (2001). The results by both formulations showed there is significant univariate global spatial autocorrelation for all four causes of CVD deaths. This suggests that there is some form of disease clustering. In turn, the local univariate and bivariate Moran's I were then used to determine hotspots and coldspots of cardiovascular deaths in South Africa.

2 Spatial Autocorrelation

Spatial autocorrelation (SAC) is the correlation of a geo-referenced variable to itself geographically. If there is geographical interdependence between geo-referenced observed values, then these data are said to exhibit spatial autocorrelation. When there are random spatial patterns then the data shows no spatial autocorrelation. Spatial autocorrelation measures the degree to which one area is similar or dissimilar to its geographically contiguous areas. Spatial autocorrelation, such as the common Pearson's autocorrelation function, can be positive or negative. Positive spatial autocorrelation occurs when geographically contiguous areas are similar, while negative spatial autocorrelation occurs when the geographically contiguous areas are not similar.

In this chapter, we tested for the presence of SAC in a given municipal neighbourhood under the null hypothesis of spatial randomness (H_0 : There is no spatial autocorrelation). The measures that are used to test the extent of spatial autocorrelation are divided into global and local measures of spatial autocorrelation. Global measures that were used to determine the presence of SAC for the whole of South Africa are univariate global Moran's I for individual CVDs, while bivariate Moran's I and Lee's L were used to determine pairwise joint global SAC. Global indicators of spatial autocorrelation (GISA) were only used to confirm if there is any form of clustering in the whole of South Africa. They do not reveal actual clusters (Waller and Gotway, 2004). Local indicators of spatial autocorrelation (LISA) are the ones that were used to reveal actual clusters at local municipal neighbourhoods of South Africa. Local measures that were used to determine the hotspots of CVDs are the univariate Moran's I that was used to detect clusters at local municipality level for individual CVDs and bivariate Moran's I that was used to detect pairwise joint clusters of the CVDs.

2.1 Spatial Weights

Knowledge of the neighbourhood structure of the regions under study is important for one to be able to quantify location in order to analyse spatial autocorrelation. The neighbourhood structure is represented as a proximity matrix known as a spatial weight matrix, W. A spatial weight matrix, $\{w_{ij}\}_{i,j=1}^{n}$, is an $n \times n$ matrix that defines the closeness or connectedness of two areas A_i and A_j in space. Spatial weight matrices can either be contiguity (neighbourhood) or distance-based. A contiguity structure shows how one area is located in relation to others, whereas distance-based structures show the relative Euclidean spatial distance of one area from the others. In contiguity structures, one would expect neighbours to have more spatial dependence than those that are far away. In distance-based neighbourhood structures, spatial dependency is expected to decline as the distance between areas increases. Areas that are far from each other should exhibit spatial heterogeneity (dissimilar relationships), while those that are close should show similar relationships.

The spatial contiguity matrices are the simplest, there are in terms of neighbourhood structure definition, and their contiguity-based spatial weights are defined as follows:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are close or connected or neighbours} \\ 0 & \text{otherwise,} \end{cases}$$

where areas A_i and A_j are said to be neighbours or connected if either: (1) they share a border (rook contiguity/simple contiguity); or (2) they share a corner (bishop contiguity); or (3) they share either a border or a corner (queen contiguity).

The simplest of the distance-based spatial matrix, like the contiguity matrices, is also a binary connectivity matrix defined such that two areas A_i and A_j are neighbours if the distance between them, d_{ij}, is less than a specified distance, say δ, beyond which autocorrelation is not expected. This structure is called the cross-hatched or distance band contiguity. Similarly defined is the k-nearest neighbour contiguity, where area A_j is one of the k areas close to A_i.

Functional distance-based spatial weight matrices have also been formulated. One such example is that based on the power function, $w_{ij} = d_{ij}^{\alpha}$, where α is the power parameter. When α is equal to negative 1, we have an inverse distance, and when it is equal to 2, we have a quadratic inverse distance that is also known as the gravity model. The distance between two areas A_i and A_j, d_{ij}, can be measured from the centroid of the areas or from major cities or any points so chosen to be representative of the areas. There are other forms of spatial contiguity not discussed here and can be found in Waller and Gotway (2004).

There are times when some areas have or are suspected to have more neighbours than others. This can occur with irregular polygons where certain areas may be smaller or bigger in size than others and thus have more neighbours than others. One

may want to adjust for this fact by creating proportional weights for the number of neighbours for an area. This is achieved through the creation of a row-standardised weight matrix whose entries will be given by

$$w_{ij}^{std} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}.$$

This standardisation is appropriate for this study in which irregularly shaped municipalities of South Africa are considered as the unit of analysis. In addition, the queen's contiguity weight matrix was preferred over distance-based weight matrices as distances between municipalities are big and describing neighbourliness for municipalities with long distances between them will not be meaningful.

3 Univariate Spatial Autocorrelation

3.1 Univariate GISA

The global Moran I statistic is the most popular of the GISAs. It measures the extent of the linear relation between the observed geo-referenced data $x_i (i = 1, 2, ..., n)$ and their corresponding spatial lags (or weighted mean values for the geographical contiguous areas) measured in terms of their deviations from their average. Here, n represents the number of samples or regions under consideration, which in this case are the 234 local municipalities of South Africa. The global Moran's I is defined by

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ij}} \cdot \frac{\sum_{i=1}^{n} z_i \sum_{j=1}^{n} \upsilon_{ij} \cdot z_j}{\sum_{i=1}^{n} z_i^2}, i \neq j, \tag{1}$$

where $\mathbf{Z}_X^T = [z_i] = [(x_i - \bar{x})/\sigma_x]$ is a vector of the standard normalised values of the x_i's and $\mathbf{V} = [\upsilon_{ij}]$ is the spatial weight matrix, which is a measure of the spatial proximity between municipality i and municipality j. This study makes use of queen contiguity spatial matrix that has been row-sum standardised (see Waller and Gotway, 2004). The advantage of standardisation is that it makes for easier interpretation. The global Moran's I can be written in matrix notation as

$$I(\mathbf{x}) = \frac{\mathbf{Z}_X^T \mathbf{V} \mathbf{Z}_X}{\mathbf{I}^T \mathbf{V} \mathbf{I}}. \tag{2}$$

The Moran's I ranges from -1, which will indicate a perfect random dispersion, to $+1$, which will indicate perfect correlation or clustering. Its interpretation is based on its theoretical expected value under the null hypothesis of spatial randomness that is given by $E[I] = \frac{-1}{n-1}$ (Griffith, 1987). When $I > E[I]$, it means that there is positive spatial autocorrelation. In other words, neighbouring areas have similar

attributes compared to further areas. If $I < E[I]$, then there is negative spatial autocorrelation. If $I = 0$, the slope is zero, then there is no spatial autocorrelation. It suffices to use the descriptive approach when making inferences about spatial autocorrelation using the global Moran's I for exploratory purposes. However, further significance tests may be done under the assumptions of normality or randomness of the processes. This study makes use of the Monte Carlo simulations that are steeped in Mantel's permutation approach (Mantel, 1967). The framework is provided in Lee (2004).

There are other global measures of univariate spatial autocorrelation such as the recently developed Lee's S statistic (Lee, 2004) and Geary's C statistic. These generally gave similar results with the univariate Moran's index and are not included in this analysis.

3.2 Univariate LISA

Having established the presence of an underlying pattern or spatial clustering in the data using global measures such as the Moran's I discussed in the section above, one may be interested in detecting hotspots of increased rates or coldspots of reduced rates that could have caused the global statistic to be significant. Furthermore, one can also identify outliers using LISA. Hotspots and coldspots are associated with positive spatial autocorrelation. Outliers are identified when the sign for local spatial autocorrelation negates that of the global spatial autocorrelation. For instance, when the global statistics are saying there is positive spatial autocorrelation, then local areas with negative spatial autocorrelations will be spatial outliers. Although there are various LISA (Anselin, 1995), this study makes use of only the most widely used method that is the local Moran I technique. The standardised local Moran's I statistic is given by

$$I_i = \frac{n^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ij}} \cdot \frac{\sum_{i=1}^{n} z_i \sum_{j=1}^{n} \upsilon_{ij} \cdot z_j}{\sum_{i=1}^{n} z_i^2}, i \neq j. \tag{3}$$

The local Moran's I can be written in matrix notation as

$$I_i(\mathbf{x}) = \frac{\mathbf{Z}_X^T \mathbf{V}_i \mathbf{Z}_X}{\mathbf{I}^T \mathbf{V} \mathbf{I}}, \tag{4}$$

where $\frac{1}{n} \cdot \sum_{i=1}^{n} I_i = I$ and $\mathbf{V}_i$ is a global spatial weight matrix whose entries are zero with the exception of the entries in the ith row. The expected value under the assumption of randomness is $E[I_i] = \frac{-1}{n-1}$, and the interpretation is as with Global Moran's I. Significance tests were done based on Monte Carlo tests based on the framework provided by Lee (2004).

4 Bivariate Spatial Autocorrelation

4.1 Bivariate GISA

Wartenberg (1985) was the first to develop the concept of a multivariate spatial correlation (MSC) measure that was based on Mantel (1967) cross-product statistic. The idea behind Wartenberg's (1985) bivariate Moran's I was to develop a bivariate simultaneously taking into consideration the univariate spatial autocorrelation of the individual variables as well as their covariance. The work of Wartenberg (1985) was extended by Anselin et al. (2002) who developed a bivariate spatial association measure for both Moran's I global and local indexes. These indexes can easily be calculated and mapped using GeoDa (Anselin et al., 2006).

Similar to the univariate Moran's I, the bivariate counterpart can also be calculated at global and local levels. Given two spatially dependent variables Y_k and X_l for a given location, then the global test statistic for the bivariate Moran's I is given by

$$I(\mathbf{xy}) = \frac{\mathbf{Z}_X^T \mathbf{V} \mathbf{Z}_Y}{\mathbf{I}^T \mathbf{V} \mathbf{I}}, \tag{5}$$

where $\mathbf{Z}_Y = \frac{(\bar{y}_i - \bar{y})}{\sigma_Y}$ and $\mathbf{Z}_X = \frac{(x_i - \bar{x})}{\sigma_X}$ follow a standard normal distribution, n is the size of observations, and $\mathbf{V}$ is the spatial weight matrix that has been row-standardised.

Lee (2001) also provided a global bivariate SAC measure denoted by L. It is defined by

$$L_{X,Y} = \frac{n}{\sum_{i=1}^{n}(\sum_{j=1}^{n} v_{ij})^2} \cdot \frac{\sum_{i=1}^{n}[(\sum_{j=1}^{n} v_{ij}(x_j - \bar{x}))(\sum_{j=1}^{n} v_{ij}(y_j - \bar{y}))]}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{6}$$

The global Lee's L can be written in matrix notation as

$$L(\mathbf{xy}) = \frac{\mathbf{Z}_X^T (\mathbf{V}^T \mathbf{V}) \mathbf{Z}_Y}{\mathbf{I}^T (\mathbf{V}^T \mathbf{V}) \mathbf{I}}. \tag{7}$$

Dray et al. (2008) suggested a method that will make the bivariate Moran's spatial association measure satisfy the conditions of Lee (2001). This was done by using a spatial weight matrix $\frac{\mathbf{W}+\mathbf{W}^T}{2}$ in the derivations of Lee (2001), where $\mathbf{W}$ is the original weight matrix. Using this transformation, Dray et al. (2008) developed a bivariate spatial association measure that we will refer to as Dray's H. This bivariate measure by Dray et al. (2008) is given by

$$H_{X,Y} = \frac{1}{2}[\sqrt{SSS_X} \cdot r_{X,\widetilde{Y}} + \sqrt{SSS_Y} \cdot r_{Y,\widetilde{X}}], \tag{8}$$

where SSS_X is the spatial smoothing scalar (Lee, 2001) and is given by

$$SSS_X = \frac{\sum_{i=1}^{n}(\widetilde{x_i} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

4.2 Bivariate LISA

The local bivariate Moran's I, which in essence is a measure of the linear relationship between an observed value at location i and the spatial lag of its neighbouring locations j (average of observed values of its neighbours), is given by

$$I_i(\mathbf{xy}) = \frac{\mathbf{Z}_X^T \mathbf{V}_i \mathbf{Z}_Y}{\mathbf{I}^T \mathbf{V} \mathbf{I}}. \tag{9}$$

5 An Application

5.1 Data

This chapter, for illustrative purposes, uses cardiovascular mortality data that has been coded using the International Statistical Classification of Diseases and Related Health Problems [ICD-10] (World Health Organization, 2004) from South Africa's vital registration system. The ICD-10 defined broad groups of causes of death (COD) data attributed to cerebrovascular heart diseases (CVAs), hypertensive heart diseases (HHDs), ischaemic heart diseases (IHDs), and diabetes (DBT) were derived for the years 2001 and 2011. In terms of nomenclature, CVAzy, IHDzy, HHDzy, and DBTzy will represent mortality due to cerebrovascular, ischaemic, hypertensive heart conditions, and diabetes in the year zy, respectively. Here, zy takes values 01 and 11, representing the years 2001 and 2011, respectively. Data were derived for the ages 30–70 years that represent an age group in which premature mortality occurs.

Table 1 shows the distribution of deaths in South Africa for the years 2001 and 2011 by the age groups 0–29, 30–70, and 71 years and over. Overall, the total deaths due to HHD increased the most from 10769 in 2001 to 15609 deaths in 2011, an increase of 44.9%. It can also be seen that DBT increased by almost the same percentage (44.2%) from 14568 deaths in 2001 to 21056 deaths. CVA deaths increased by 14.6% (from 22590 to 25983), while IHD increased by only 2.1% (from 11779 to 12023) over the same period. In the age group 30–70 years, Table 1 shows that there has been a slight decrease in the number of deaths for CVA (−0.4%) and IHD (−5.8%) between 2001 and 2011, while HHD increased by

about 22.3%. It is in this 30–70-year age group that premature mortality needs to be reduced and analysis will be done for this age group.

The data quality issues associated with DNF data include, among others, garbage codes, misclassification, and incompleteness of death registration (Joubert et al., 2013; Pillay-van Wyk et al., 2011). Adjustments have to be made to these data to minimise bias that may be attributed to these quality issues. Correcting the rate of mortality usually involves using the age- or sex-specific death rates of standard population to which the mortality rates of interest are adjusted (Birnbaum et al., 2011). There are two problems with this approach. First, the choice of standard population to use is usually arbitrary and subjective (Birnbaum et al., 2011). Second, the standardised mortality rates assume that the characteristics of small and large areas are the same and the resulting estimates have been criticised for not being representative enough of the geographic distribution of the rates (Clayton and Kaldor, 1987; Sarndal, 1984). Thus, alternative techniques have been sought to estimate rates at a local level for compromised data. These techniques are briefly described in the next subsection.

5.2 *Statistics Methods in Rate Estimation*

The EB approach and the Poisson regression model were considered for estimating the mortality rates at municipal level. In the EB approach, the number of observed deaths in municipality i, and due to disease j, O_{ij} is allowed to follow a Poisson distribution with both the mean and the variance equal to the product of P_i, the population at risk in municipality i, and π_{ij}, the unknown underlying risk of mortality due to disease j in municipality i. It follows that the observed deaths are conditioned on the varying underlying risk of mortality, and we write

$$O_{ij}|\pi_{ij} \sim Poisson(\pi_{ij} P_i). \tag{10}$$

Additionally, the mortality risk, π_{ij}, is allowed to follow a Gamma distribution with shape parameter α and scale parameter ϕ. That is

$$\pi_{ij} \sim Gamma(\alpha, \phi), \tag{11}$$

where $E(\pi_{ij}) = \frac{\alpha}{\phi}$ and $Var(\pi_{ij}) = \frac{\alpha}{\phi^2}$. According to Bayes theorem, the following proportionality holds:

$$Pr(\pi_{ij}|O_{ij}) \propto Pr(O_{ij}|\pi_{ij}) \times Pr(\pi_{ij}), \tag{12}$$

and, importantly, the conditional posterior also follows a Gamma distribution with shape parameter $\alpha + O_{ij}$ and scale parameter $P_i + \phi$. It follows that

Table 1 Distribution of the number of deaths across age groups by year, South Africa

Year	Age group	CVA		HHD		IHD		DBT	
		Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
2011	0–29	485	1.9%	148	0.9%	165	1.4%	271	1.3%
	30–70	12,196	47.1%	7180	46.0%	6183	51.4%	12,063	57.3%
	71+	11,946	46.1%	7561	48.4%	5248	43.6%	7736	36.7%
	Missing	1266	4.9%	720	4.6%	427	3.6%	986	4.9%
	Total	25,893	100.0%	15,609	100.0%	12,023	100.0%	21,056	100.0%
2001	0–29	593	2.6%	159	1.50%	129	1.1%	250	1.7%
	30–70	12,241	54.2%	5873	54.50%	6564	55.7%	9185	62.9%
	71+	9756	43.2%	4735	44.00%	5074	43.1%	5145	35.3%
	Missing	0	0.0%	2	0.00%	12	0.1%	17	0.1%
	Total	22,590	100.0%	10,769	100.00%	11,779	100.0%	14,597	100.0%

Key: DBT, Diabetes; CVA, Cerebrovascular heart disease; HHD, Hypertensive heart disease; IHD, Ischaemic heart disease

$$\pi_{ij}|O_{ij} \sim Gamma(\alpha + O_{ij}, P_i + \phi). \tag{13}$$

Since $E(\pi_{ij}|O_{ij}) = \frac{O_{ij}\alpha}{P_i+\phi}$, it can be deduced that the raw rates, $\widehat{\pi_{ij}} = \frac{O_{ij}}{P_i}$, can be adjusted using posterior distribution, $Pr(\pi_{ij}|O_{ij})$, if α and ϕ can be derived from the prior distribution, $Pr(\pi_{ij})$. In fact, it can be shown that the EB estimate of the underlying mortality is the expected value of the distribution of the conditional posterior:

$$\widehat{\pi_{ij}}^{EB} = E(\pi_{ij}|O_{ij}) = \frac{O_{ij}\alpha}{P_i + \phi}, \tag{14}$$

where the parameters α and ϕ are determined from the observed data.

In the Poisson regression approach, the expected mean of $O_{ij}(= \pi_{ij}P_i)$, denoted by μij (the expected number of deaths in municipality i dying a premature death (between 30 and 70 years) for a given disease, j), is modelled as

$$\begin{aligned}\mu ij &= \log(\pi_{ij}P_i) = log(P_i) + \alpha + \beta_1(p_{age}) + \beta_2(p_{race}) + \beta_3(poverty)(\pi_{ij}|O_{ij}) \\ &= \frac{O_{ij}\alpha}{P_i + \phi} + \varepsilon_{ij},\end{aligned} \tag{15}$$

where p_{age} is the proportion of the age group 30–70 that are aged 50 to 70 in the population of municipality i, p_{race} is the proportion of a given race in municipality i for the given age group, and $poverty$ is the level of poverty in municipality i measured by the official South African multidimensional poverty index obtained from the 2001 and 2011 census data (Statistics South Africa, 2014b).

A descriptive summary of the raw, smoothed, and adjusted rates for the age group 30–70 years of interest to this study is given in Table 2. Generally, we have mean rates of the same order for all the three rates. The major difference in the rates, however, is found in the ranges, where observed raw rates have the highest range in all instances owing to very high maximum values. Further investigations revealed that the municipalities with the smallest populations are also the ones with the highest (as well as smallest) mortality rates. The raw rates are sensitive to small population counts, resulting in instability. This is a well-documented problem when using raw mortality rates. Empirical Bayes rates are known to alleviate this problem (Leyland and Davies, 2005; Marshall, 1991). Adjusting for covariates also managed to alleviate the problem by reducing the maximum values and increasing the minimum values of the observed rates.

6 Results

The raw, smoothed, and adjusted mortality rates at municipal level for each of the four disease conditions studied in South Africa were mapped on a choropleth map. The results are shown in Figs. 1 and 2 for the years 2001 and 2011, respectively.

Table 2 Descriptive statistics of raw, EB smoothed, and adjusted mortality rates across municipalities for CVA, IHD, HHD, and DBT, South Africa

Model	Mean	SD	Minimum	Maximum
CVA 2001				
Adj	86.37	18.96	41.26	157.10
EB	85.96	47.25	3.02	296.56
RR	88.75	59.63	0.00	389.89
CVA 2011				
Adj	72.90	17.33	37.08	151.76
EB	76.24	48.18	7.20	377.59
RR	79.66	60.55	0.00	477.90
IHD 2001				
Adj	48.81	33.25	6.63	147.98
EB	43.11	37.79	1.49	362.80
RR	44.65	45.24	0.00	407.68
IHD 2011				
Adj	37.48	20.76	10.71	116.24
EB	34.31	28.91	2.09	259.34
RR	36.32	38.41	0.00	309.69
HHD 2001				
Adj	43.56	183.69	12.96	149.08
EB	38.68	27.52	1.77	226.84
RR	38.69	33.72	0.00	244.37
HHD 2011				
Adj	44.36	17.34	12.32	129.73
EB	44.01	29.33	5.37	164.24
RR	45.89	36.51	0.00	180.63
DBT 2001				
Adj	64.72	18.04	21.70	122.78
EB	53.06	39.04	2.32	403.50
RR	51.66	46.52	0.00	457.39
DBT 2011				
Adj	71.78	17.76	29.68	132.86
EB	66.49	43.99	7.73	353.56
RR	67.06	53.37	0.00	405.13

Key: SD = Standard deviation

High-mortality-risk municipalities are indicated by the darkest colour (quartile 4 or upper quartile). Municipalities with low mortality rate risks are shown by a hollow (quartile 1 or lower quartile). The darker the colour the higher the risk of mortality due to each condition.

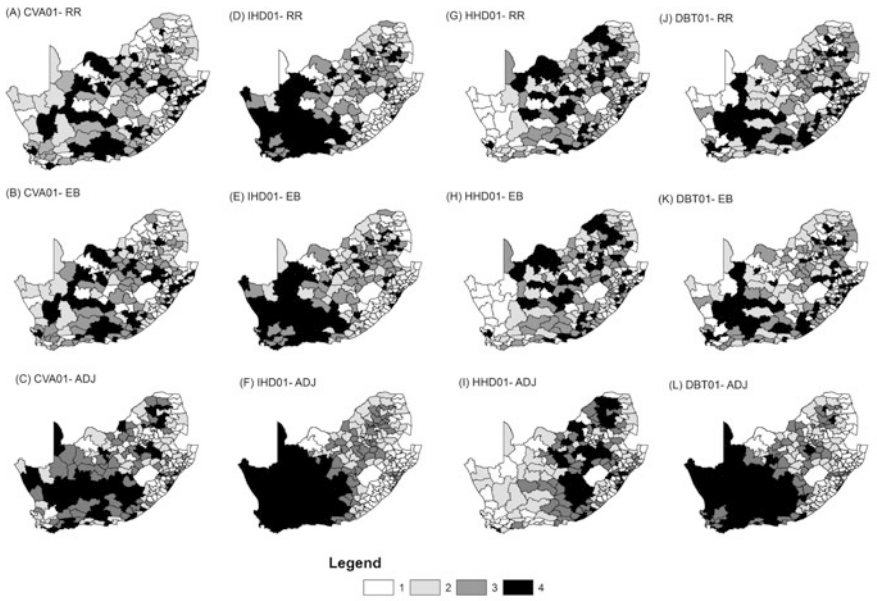


Fig. 1 Choropleth maps showing the distribution of raw, smoothed, and adjusted mortality rates for the year 2001

Generally, the quantile maps reveal some form of clustering. Consider Fig. 1A–C to see the effects of smoothing and adjustment for covariates. The quantile map of the observed raw rate (CVA01-RR) in Fig. 1A and the smoothed rate (CVA01-EB) in Fig. 1B are almost similar in terms of their spatial distributions. There is not much difference between the distribution of mortality rates before and after smoothing. It seems that the effects of stabilising the crude rates with the EB approach have not, based on the evidence of the choropleth maps, improved the ability to discern areas of higher mortality risk.

Adjusting for covariates, as the case of CVA01-Adj in Fig. 1C, results in a more defined cluster in the south-west part of the country when compared with raw and smoothed rates in Fig. 1A–B. This is the general pattern with all the other disease conditions, with dark colours more noticeable for adjusted rates than for raw and smoothed rates, and are mostly concentrated in the western part of the country. Only HHD clustering seems to stretch from the middle of the country toward the eastern part of the country. The spatial patterns exhibited in Fig. 1 for the year 2001 are similar to the spatial patterns exhibited by the corresponding mortality rates in Fig. 2 for the year 2011. In the next section, the statistical significance tests of spatial autocorrelations were done and discerned clusters mapped.

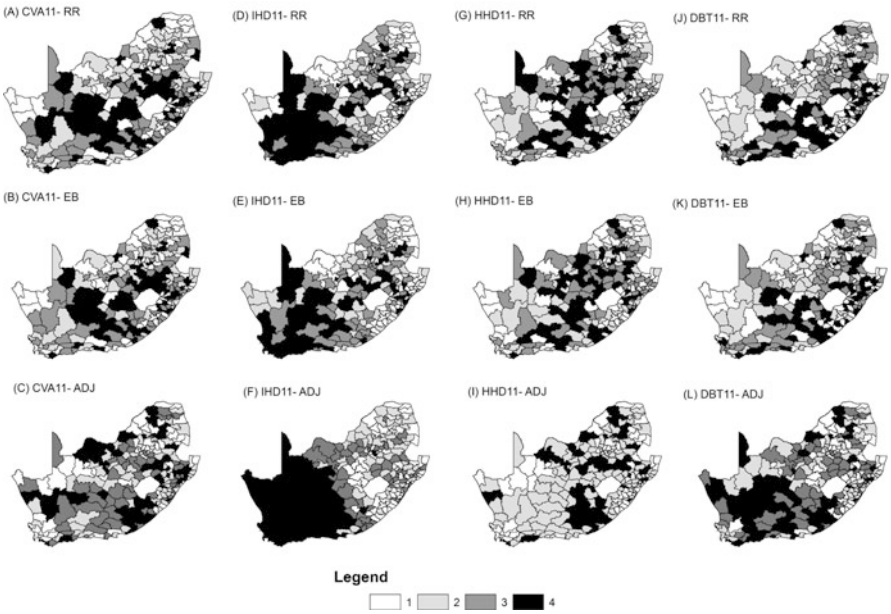


Fig. 2 Choropleth maps showing the distribution of raw, smoothed, and adjusted mortality rates for the year 2011

6.1 Univariate Cluster Analysis

6.1.1 Univariate Global Spatial Autocorrelation

The previous assessment of geographical variations for raw, smoothed, and adjusted rates has shown evidence of clustering in CVD and diabetes outcomes. In order to formally investigate spatial association, we measured the association in a formal way by using univariate (in this section) and bivariate clustering statistics (see subsection 4.3 for details). Table 3 presents the derived values for each CVD for the whole of South Africa for the years 2001 and 2011. For comparison purposes, the derivations were done using raw, smoothed, and adjusted rates.

The univariate Moran's I test in Table 3 confirms that the distribution of the four conditions IHD, CVA, DBT, and IHD varies geographically, when adjusted rates are used (p-value < 0.05). Both raw and smoothed rates failed to detect clusters of DBT, while the raw rate further failed to detect any significant clustering for CVA01 (p-value >0.05). The geographic variation based on adjusted rate is significant for both the years 2001 and 2011. In all the cases, the calculated statistics for Moran's I are all positive and significant across the years. This means that the likelihood of the spatial patterns generated by mortality due to each of the three CVDs being due to random chance is negligibly small (less than 5%). Thus, one can conclude that the probability is high that municipalities that are nearer to each other tend to have

Table 3 Univariate global Moran's spatial autocorrelations for the model residuals, raw, smoothed, and adjusted mortality rates due to CVA, IHD, DBT, and HHD in 2001 and 2011

Model	Moran's I (Estimates)	p-value	Moran's I (Residuals)	p-value
CVA 2001				
Adj	0.422	<0.001	0.038	†
EB	0.021	<0.001	0.068	<0.05
RR	0.029	†		
CVA 2011				
Adj	0.297	<0.001	0.109	<0.05
EB	0.078	<0.05	0.121	<0.05
RR	0.088	<0.05		
IHD 2001				
Adj	0.849	<0.001	−0.003	†
EB	0.108	<0.001	0.318	<0.05
RR	0.251	<0.001		
IHD 2011				
Adj	0.821	<0.001	0.149	<0.05
EB	0.093	<0.001	0.150	<0.05
RR	0.176	<0.05		
HHD 2001				
Adj	0.445	<0.001	0.066	†
EB	0.144	<0.001	0.045	†
RR	0.218	<0.001		
HHD 2011				
Adj	0.329	<0.001	0.112	<0.05
EB	0.135	<0.001	0.054	†
RR	0.101	<0.05		
DBT 2001				
Adj	0.684	<0.001	0.063	†
EB	0.005	†	0.003	†
RR	0.006	†		
DBT 2011				
Adj	0.316	<0.001	0.064	†
EB	0.038	†	0.003	†
RR	0.030	†		

Key: † = Insignificant p-values

comparable baseline mortality rates than the distant municipalities. In other words, there is some form of clustering exhibited by all three CVDs at the 5% significance level. This is a reflection of what is seen in the quantile maps in Figs. 1 and 2.

Table 3 also shows the residuals of the smoothed and adjusted rates for each of the CVDs for the years 2001 and 2011. The statistical autocorrelation of the residuals, based on the Moran's index, was found to be insignificant for some of the fitted models for adjusted rates (CVA11, IHD11, and HHD11) and the smoothed rates

(CVA01, HHD01, CVA11, and IHD11). This statistical autocorrelation analysis of the residuals is not a criterion for diagnostic checks for generalised linear models or EB approach, but it would be preferable if residual spatial autocorrelations were not significant. This is because the presence of spatial autocorrelations in the residuals suggests that the model is not adequately specified. That is to say there may exist some unmeasured covariates not specified in the model that may help in explaining the variation of mortality rates across the municipalities. Introducing spatial random effects or an eigenvector spatial filter (Griffith and Chun, 2014) did not remove the residual spatial autocorrelations, so the original specified Poisson regression model with covariates only was returned using the rule of parsimony. In the next section, we looked at the LISA maps for all rates for visual comparison purposes only, irrespective of whether they are significant or not.

6.2 Univariate "Hot-Spot" Analysis

Having established the presence of clustering using GISA statistics in the previous section, LISA maps were used to determine the local municipality level clusters for raw, smoothed, and adjusted rates for CVA, IHD, DBT, and HHD. The results are shown in Figs. 3 and 4 for the years 2001 and 2011, respectively. "Hot-spots," which are municipalities of high mortality incidences that are surrounded by municipalities with high mortality incidences, are indicated by a "High-High" (H-H) key on the map, while the "cold-spots", which are municipalities of low mortality incidences that are surrounded by municipalities with low mortality incidences, are indicated by a "Low-Low" (L-L) key. In addition, there are outliers indicated by "High-Low" (H-L), which are municipalities of high mortality incidences that are surrounded by municipalities with low mortality incidences, and "Low-High" (L-H), which are municipalities of low mortality incidences that are surrounded by municipalities with high mortality incidences. Municipalities whose clustering is not significant are denoted by "Not Significant" (N-S) key and have a white (hollow) shade. The "hot-spots" have a black shade in the map, while "cold-spots" have a white colour.

Adjusted rates in Fig. 3 have noticeable and well-defined clusters as compared to raw and smoothed rates. Generally, clusters are found in the south-west part of the country, except for HHD that has clusters in the south- and north-east part of the country. The clusters for CVA and DBT seem to have reduced in size over the 10-year period under review. In Fig. 3C, for example, CVA01 LISA-derived clusters comprise 31 municipalities, but these have been more than halved to 16 municipalities in 2011 (see Fig. 4C). The disappearance or movement of the cluster from the south-east may be due to intervention programmes aimed at alleviating the problem in the area. However, further investigation may help to explain what is truly happening, especially with DBT whose data suggest that mortality due to this disease has increased over the 10-year period under review.

The adjusted rate-based clusters for IHD (see Figs. 3F and 4F) and HHD (see Figs. 3I and 4I) have not changed much over the period. This shows that the spatial

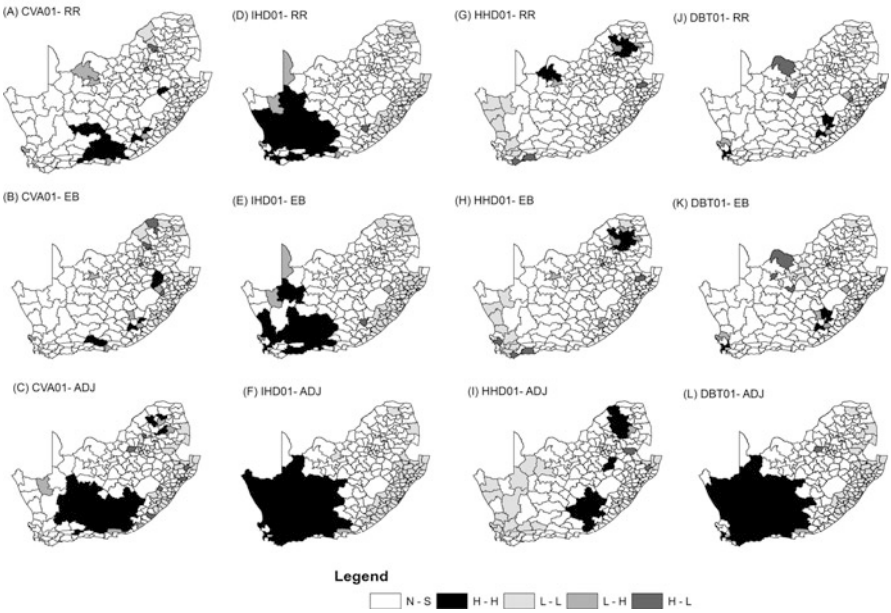


Fig. 3 Univariate Moran's I LISA maps showing the distribution of clusters of raw, smoothed, and adjusted mortality rates for the year 2001

dynamics of IHD and HHD are stable over the period under study, with IHD "hot-spots" located in the centre and spanning all the way to the south-west coast of the country. The LISA analysis for HHD (Figs. 3I and 4I) reveals two clusters in the south- and north-east part of the country for both the years 2001 and 2011.

6.3 Bivariate Analysis

6.3.1 Bivariate Association of Individual CVD Maps Over Time

It was shown in Table 3 that the geographical variation of mortality due to each disease is significant over the years and the clusters have been shown in Figs. 3 and 4. It remains to be seen if spatial patterns of each CVD risk are significantly different for the two time periods under review. These differences or similarities of the spatial distribution of each disease outcome for the years 2001 and 2011 were established here using bivariate spatial autocorrelation measures. This will happen to ascertain any changes in the spatial patterns of individual health outcomes for the 10-year period under review. Bivariate Lee's L, Moran's I, and Dray's H were used to determine if there is spatial dependence between the health outcome data for the years 2001 and 2011. Results are shown in Table 4.

Table 4 Bivariate global spatial autocorrelations measuring spatial dependence of individual CVD rates for raw, smoothed, and adjusted mortality rates between the years 2001 and 2011

Association (X-Y)		Lee 2001 ($L_{X,Y}$)			Anselin 2002 ($I_{X,Y}$)			Dray 2008 ($H_{X,Y}$)		
X	Y	RR	EB	Adj	RR	EB	Adj	RR	EB	Adj
CVA01	CVA11	0.115**	0.125**	0.222***	0.060**	0.075**	0.179***	0.065**	0.075**	0.185***
IHD01	IHD11	0.250***	0.259***	0.811***	0.239**	0.243**	0.819***	0.254***	0.252***	0.821***
DBT01	DBT11	0.065†	0.076**	0.367***	0.065**	0.079**	0.359***	0.058**	0.071**	0.379***
HHD01	HHD11	0.084**	0.097**	0.331***	0.082**	0.094**	0.322***	0.084**	0.085**	0.325***

Key: *** = p-values <0.001; **= p-values < 0.05; † = Insignificant p-values

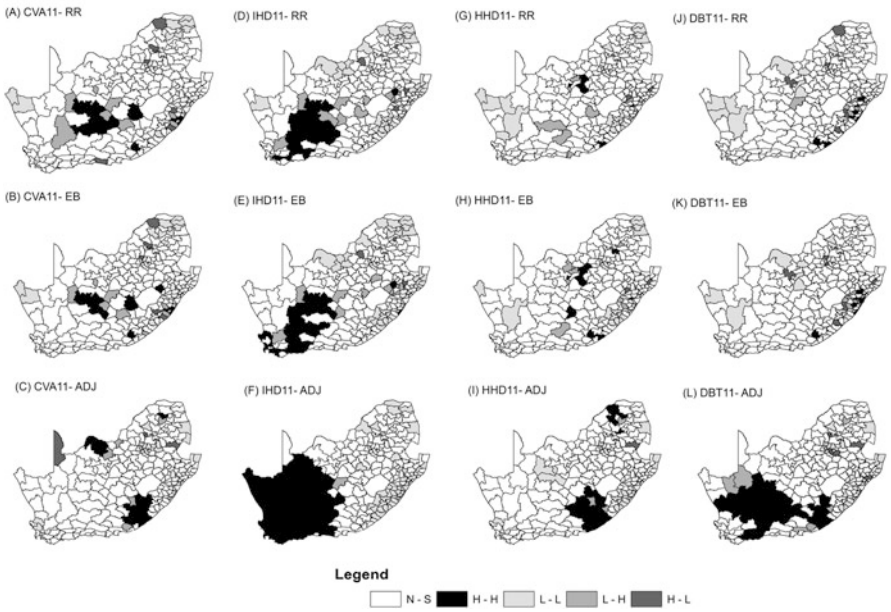


Fig. 4 Univariate Moran's *I* LISA maps showing the distribution of clusters of raw, smoothed, and adjusted mortality rates for the year 2011

All the three indicators of bivariate spatial autocorrelation applied to these data reveal evidence of spatial dependence for all four health outcomes on how each outcome is spatially distributed for the years 2001 and 2011. It can, thus, be concluded that the spatial distribution of the risk of mortality due to each CVD has not significantly changed over the course of the 10-year period under review. As expected, the bivariate Moran's *I* and Dray's *H* show similar results. This is not surprising as the methods are based on the same derivation. The bivariate LISA analyses of the combinations in Table 4 derived from the raw, smoothed, and adjusted rates are presented in Fig. 5.

In Fig. 5F, as an example, observed "hot-spots" are areas of high mortality of IHD in 2001 whose neighbourhood in 2011 also exhibits high mortality of IHD to form a co-cluster of high mortality for the two time points in the south-western part of the country. The CVA and HHD co-clusters are similar and are found in the south- and north-east part of the country.

6.3.2 Bivariate Spatial Association Between Two CVDs at a Point in Time

We also looked at determining spatial dependency between two different CVDs at a cross-section. One can hypothesise that CVDs should co-cluster or show spatial dependency at a point in time as they share risk factors. Table 5 presents the bivariate association measure values calculated for the possible combinations of the three

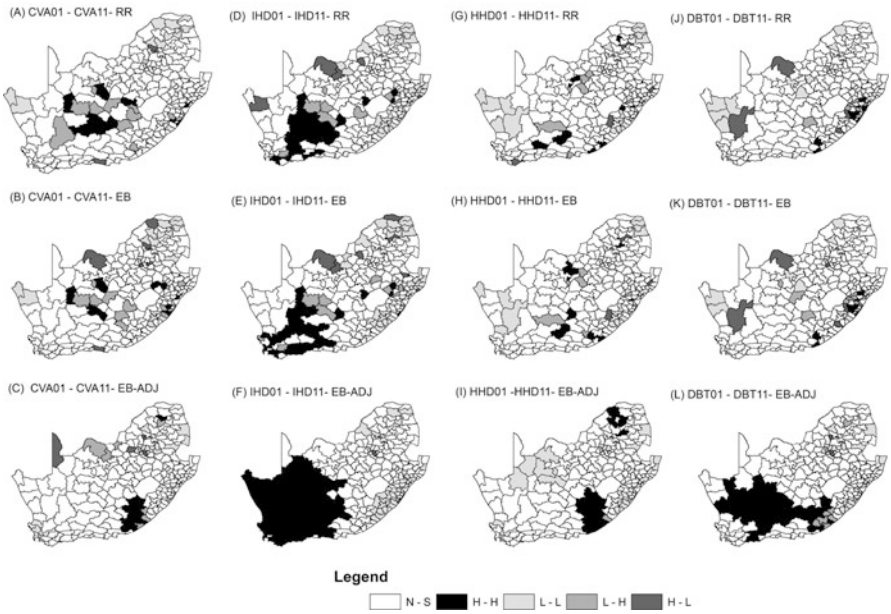


Fig. 5 The raw, smoothed, and adjusted mortality rate-based bivariate Moran's I LISA maps between same CVDs for the years 2001 and 2011

CVDs for the years 2001 and 2011 to determine spatial dependence based on raw, smoothed, and adjusted rates data.

The bivariate Moran's I and Dray's H once again showed similar results. All three methods generally agree, based on the raw and smoothed rates, that there is no evidence of spatial dependence between all the associations tested. However, adjusted rate-based tests revealed significant spatial dependence between the following maps: CVA01 and IHD01; CVA01 and HHD01; and CVA11 and HHD11. Importantly, DBT was found to have a significant association with all the three CVDs. This is in line with expectation as DBT is a well-known biomarker for CVDs. The other associations were either insignificant or their associations were purely random with a negative Moran's index. The significant joint local "hot-spots" of the CVD associations, based on adjusted rate data, are shown in Fig. 6.

Focusing on the most recent 2011 data, it can be seen that the "hot-spots" of DBT and the three CVDs in Fig. 6F–H are located in the south-west part of the country. The joint clusters of CVA and HHD for the year 2011 are in the south- and north-west of the country. The joint clusters of CVA-DBT and IHD-DBT have reduced in size over the period under review. This may be attributable to intervention programmes. However, joint clusters of HHD and DBT that were not in existence in 2001 have formed over the period under review as both the deaths and crude national rates attributable to the two diseases have increased over the period as was shown in Tables 1 and 2.

Table 5 Bivariate global spatial autocorrelations measuring spatial dependence of individual CVD rates between two time periods for raw, smoothed, and adjusted mortality rates between the years 2001 and 2011

Association (X-Y)		Lee 2001 ($L_{X,Y}$)			Anselin 2002 ($I_{X,Y}$)			Dray 2008 ($H_{X,Y}$)		
X	Y	RR	EB	Adj	RR	EB	Adj	RR	EB	Adj
CVA01	IHD01	0.159†	0.165**	0.338***	0.014†	−0000†	0.344***	0.018†	0.003†	0.359***
CVA11	IHD11	0.173†	0.166†	−0.084†	0.067**	0.051†	−0.105†	0.066**	0.048†	−0.113†
CVA01	HHD01	0.046†	0.046†	0.295***	−0.070**	−0.075**	0.279***	−0.066†	−0.073†	0.281***
CVA11	HHD11	0.135†	0.139†	0.350***	0.017†	0.012†	0.294***	0.011†	0.007†	0.274***
CVA01	DBT01	0.154†	0.176†	0.436***	−0.040†	−0.037†	0.432***	−0.033†	−0.032†	0.442***
CVA11	DBT11	0.187†	0.196†	0.269***	0.009†	−0.042†	0.196***	0.007†	0.013†	0.196***
DBT01	IHD01	0.192†	0.203**	0.713***	0.002†	0.003†	0.719***	0.002†	0.002†	0.725***
DBT11	IHD11	0.108†	0.117†	0.302***	0.008†	0.014†	0.289***	0.009†	0.015†	0.312***
HHD01	IHD01	−0.035†	−0.018†	−0.116†	−0.100**	−0.100**	−0.121†	−0.107**	−0.095†	−0.128†
HHD11	IHD11	0.105†	0.097†	−0.089†	0.034†	0.018†	−0.110†	0.033†	0.018†	−0.110†
HHD01	DBT01	0.042†	0.044†	0.065†	−0.069**	−0.067†	0.057†	−0.069†	−0.069**	0.058†
HHD11	DBT11	0.131†	0.134†	0.263***	−0.012†	−0.012†	0.203***	−0.011†	−0.012†	0.205***

Key: *** = p-values < 0.001; **= p-values < 0.05; † = Insignificant p-values

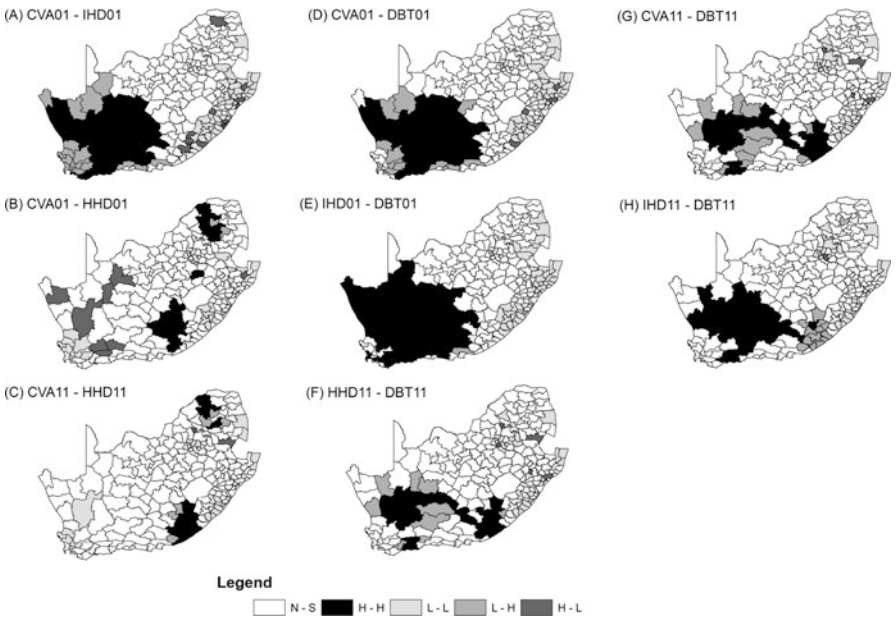


Fig. 6 The significant adjusted mortality rate-based bivariate Moran's I LISA map between two CVDs at a point in time, 2001 and 2011

7 Discussion

In this chapter, we have shown how spatial autocorrelation measures can be used to determine spatial dependence in health outcomes in South Africa. Global univariate spatial autocorrelation was used to determine the presence of spatial patterns in an individual health outcome, followed by the use of local univariate spatial autocorrelation analysis to determine local municipality clusters of high or low risk of the outcome. Global and local bivariate spatial autocorrelation were then used to determine the pairwise co-clustering of all the health outcomes across municipalities. Determination of joint clusters is more efficient than univariate cluster analysis since cardiovascular diseases and their risk factors are known to have similar aetiology (Kandala et al., 2013). Pairwise co-clustering has the advantage of providing us with more insight into how multiple interrelated health outcomes interact in space, which is more than what individual spatial autocorrelation can reveal.

The mortality rates of each of the four diseases studied revealed statistically significant spatial clustering based on the global univariate Moran's index of autocorrelation for raw rates, EB smoothed rates, and Poisson regression-adjusted rates. However, the residuals of EB rates have significant spatial autocorrelation for four out of eight health outcomes (IHD01, IHD11, CVA01, and CVA11) with

significant level greater than 0.05 compared to two (IHD11 and HHD11) for Poisson regression-adjusted rates. This demonstrates that formal modelling using Poisson regression-adjusted rates has improved the explanation of the spatial clustering for most of the cases. Thus our discussion will focus on results involving adjusted rates.

Three global bivariate spatial autocorrelation measures, namely the original Moran's I, Lee's L, and Dray's H (Anselin et al., 2002; Lee, 2001; Dray et al., 2008), were used to detect the presence of pairwise co-clustering. They generally gave similar results and may be used to complement each other, an observation also made in Darikwa et al. (2019). It is not surprising that the results by the original and Dray's H are almost identical, since the latter is just a variant of the former (Darikwa et al., 2019).

It was shown that the spatial patterns of each CVD risk were not significantly different for the two time periods under review for all four diseases. These similarities of the spatial distribution of each disease outcome for the years 2001 and 2011 were established using bivariate spatial autocorrelation measures. This shows that the spatial distribution of the risk of mortality due to each CVD did not change significantly between the years 2001 and 2011. Therefore, since the spatial distribution of the health has been stable over the years, it gives us confidence that any identified clusters of high risk mortality will still be in existence over time. Hence, one can plan for interventions in the problematic municipalities with increased confidence. Focusing on the adjusted data, the "hot-spots" pairwise co-clusters for CVA and HHD are similar for the 10-year period and are located in the southern part of the country and north-east of the country. "Hot-spots" co-clusters for the year 2001 and 2011 data for both IHD and DBT stretch from the centre of the country westwards. "Hot-spots" co-clusters, in this case, are municipalities of high mortality rates for a given health outcome in 2011 that are surrounded by municipalities of high mortality rates of the same health outcome in 2001.

Pairwise co-clustering between the different health outcomes for a given year were found statistically significant for the following using adjusted mortality rates: CVA01-IHD01; CVA01-HHD01; CVA11-HHD11; CVA01-DBT01; CVA11-DBT11; IHD01-DBT01; IHD11-DBT11; HHD11-DBT11. There were no significant joint clusters between HHD and IHD for the years 2001 and 2011. The CVA-IHD and CVA-DBT "hot-spots" clusters are the same in 2001, but the co-clusters of the former have disappeared (not significant) in 2011, while that of the latter has shrunk in size significantly in 2011. This may be attributed to the fact that the univariate "hot-spots" clusters for CVA have also significantly reduced in size between 2001 and 2011. The co-clusters between of CVA and HHD are similar for the years 2001 and 2011 with "hot-spot" co-clusters found in the south of the country and in the north-east of the country, in line with the spatial patterns by the individual CVD conditions for the same period. Joint "hot-spots" clusters for IHD and DBT are in the south-west part of the country for this period and so are HHD and DBT clusters for the year 2011. These findings are in line with research in South Africa that has established the south-western parts of the country as problematic in as far as cardiovascular diseases are concerned (see Statistics South Africa, 2009, 2014a; Groenewald et al., 2014; Kandala et al., 2014). "Cold-spots"

clusters are also important in as far as establishing factors that are protective of cardiovascular mortality. Generally, the "cold-spots" clusters for CVDs and diabetes shown in the maps are found in municipalities that on the belt along the south-eastern boundaries of the country and in the north-east municipalities of the country. These municipalities are mostly poor rural areas.

Previous research in South Africa have used data that are age-standardised, age–sex-standardised rates, and Empirical Bayes smoothed rates (Darikwa et al., 2019; Darikwa and Manda, 2020). But there are other covariates besides age and sex that help to explain spatial distribution of health outcomes (Chien et al., 2018). Generalised linear models have been found useful in estimating prevalence/rates at local level while controlling for additional covariates (Chien et al., 2018). Just like the global empirical Bayes approach, they can improve on raw rates that may show instability in areas of small populations. Unlike the global empirical Bayes, they adjust for additional covariates and can also cater for spatial autocorrelations. In this chapter, the Poisson regression was used to account for known municipal level risk factors of CVDs, namely age, race, and poverty. This helps to improve estimates of the mortality rates by factoring in more explanatory variables. The Poisson-adjusted rates managed to detect co-clustering where the EB smoothed rates and raw rates failed, especially regarding spatial dependency of diabetes and the other three CVDs. Co-clusters developed in this study may be viewed as a first approximation that may be improved on as more information on additional covariates are incorporated and different estimation models are developed/considered and implemented.

This chapter has explored the use of multivariate spatial autocorrelations in identifying co-clusters of high-risk health outcomes. Identified co-clusters may be used as target areas in prioritising limited resources in the fight against problematic interrelated health outcomes. Identified co-cluster means that are efficiently utilised as two or more health outcomes are targeted simultaneously, instead of one at a time. Such an integrated approach will ensure resource allocation is optimised while fighting multiple diseases at the same time, which may result in minimised costs.

References

Anselin, L. (1995). Local indicators of spatial association (LISA). *Geographical Analysis, 27*, 93–115.

Anselin, L., Syabri, I. & Smirnov, O. (2002). Visualizing multivariate spatial correlation with dynamically linked windows. *Urbana, 51*, 61801.

Anselin, L., Syabri, I. & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis, 38*, 5–22.

Birnbaum, J., Murray, C., & Lozano, R. (2011). Exposing misclassified HIV/AIDs deaths in South Africa. *Bulletin of the World Health Organization, 89*, 278–285.

Chien, L., Lin, G., Li, X. & Zhang, X. (2018). Disparity of imputed data from small area estimate approaches–a case study on diabetes prevalence at the county level in the US. *Data Science Journal, 17*, 1–11.

Clayton, D. & Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics, 43*(3), 671–681.

Darikwa, T. B., & Manda, S. (2020). Spatial co-clustering of cardiovascular diseases and select risk factors among adults in South Africa. *International Journal of Environmental Research and Public Health, 17*, 3583.

Darikwa, T. B., Manda, S., & Leasoana, M. (2019). Assessing joint spatial autocorrelations between mortality rates due to cardiovascular conditions in South Africa. *Geospatial Health, 14*, 294–305.

Dray, S., Said, S., & Debias, F. (2008). Spatial ordination of vegetation data using a generalisation of Wartenberg's multivariate spatial correlation. *Ecological Modelling Journal of Vegetation Science, 19*, 45–56.

Griffith, D. A. (1987). Spatial autocorrelation: A primer. *Association of American Geographers, Resource Publications in Geography*.

Griffith, D., & Chun, Y. (2014). Spatial autocorrelation and eigenvector spatial filtering. In M. Fischer & P. Nijkamp (Eds.). *Handbook of Regional Science*. Berlin: Springer.

Groenewald, P., Bradshaw, D., Day, C., & Laubscher, R. (2014). Burden of disease. In N. Massyn, C. Day, N. Peer, A. Padarath, P. Barron & R. English (Eds.). *District Health Barometer: 2013-14*. Durban: Health Systems Trust.

Joubert, J., Rao, C., Bradshaw, D., Vos, T., & Lopez, A. (2013). Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. *PLoS One, 8*, e64592.

Kandala, N.-B., Tigbe, W., Manda, S. O. M., & Stranges, S. (2013). Geographic variation of hypertension in Sub-Saharan Africa: A case study of South Africa. *American Journal of Hypertension, 26*, 382–391.

Kandala, N.-B., Mandad, S. O., Tigbea, W. W., Mwambi, H. & Stranges, S. (2014). Geographic distribution of cardiovascular comorbidities in South Africa: a national cross-sectional analysis. *Journal of Applied Statistics, 41*, 1203–1216.

Khamis, F. G. (2012). The relationship between spatial patterns of illnesses and unemployment in Iraq-2007. *Global Journal of Health Science, 4*, 192–203.

Khamis, F. G., El-Refae, G. A., & Fares, A.-R. F. (2014). Multivariate spatial association between mortality, unemployment, divorce, and crime in Jordan-2011. *Journal of Statistical and Econometric Methods, 3*, 49–74.

Lee, S.-I. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems, 3*, 369–385.

Lee, S.-I. (2004). A generalized significance testing method for global measures of spatial association: an extension of the Mantel test. *Environment and Planning, 36*, 1687–1703.

Leyland, A. H., & Davies, C. A. (2005). Empirical Bayes methods for disease mapping. *Statistical Methods in Medical Research, 14*, 17–34.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27*, 209–220.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical bayes estimators. *Journal of the Royal Statistical Society, 40*(2), 283–294.

Matkan, A., Mohaymany, A., Shahri, M., & Mirbagheri, B. (2013). Detecting the spatial and temporal autocorrelation among crash frequencies in urban areas. *Canadian Journal of Civil Engineering, 40*, 195–203.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*, 17–23.

Pillay-van Wyk, V., Bradshaw, D., Groenewald, P., & Laubscher, R. (2011). Improving the quality of medical certification of cause of death: The time is now! *South African Medical Journal, 101*, 171–177.

Sarndal, C. E. (1984). Design-consistent versus model-dependent estimation for small domains. *Journal of the American Statistical Association, 79*, 624–631.

Statistics South Africa (2009). Mortality and causes of death in South Africa, 2007: findings from death notification. In *Statistical Release P0309.3*. Pretoria: Statistics South Africa.

Statistics South Africa (2014a). Mortality and causes of death in South Africa, 2011: findings from death notification. In *Statistical Release P0309.3*. Pretoria: Statistics South Africa.

Statistics South Africa (2014b). The South African MPI: Creating a multidimensional poverty index using census data. *Report no.: 03-10-08*. Pretoria: Statistics South Africa.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. New Jersey: Wiley.

Wartenberg, D. (1985). Multivariate spatial correlation: A method for exploratory geographical analysis. *Geographical Analysis, 17*, 263–283.

White, M. E., Schukken, Y. H., & Tanksley, B. (1989). Space-time clustering of, and risk factors for, farmer-diagnosed winter dysentery in dairy cattle. *Canadian Veterinary Journal, 30*, 948–951.

World Health Organization (2004). ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth revision. *Technical report*, W.

Zhang, C., Luo, L., Xu, W., & Ledwith, V. (2008). Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of The Total Environment, 398*, 212–221.

Bivariate Copula-Based Spatial Modelling of Health Care Utilisation in Malawi

Ellen Gondwe, Michael G. Chipeta, and Lawrence Kazembe

Abstract In health care utilisation, multiple correlated data are common. Quantifying dependencies among interrelated variables is an important statistical problem, particularly to account for the nature of their association. Copula models permit a flexible approach to modelling dependence in interrelated outcome variables. However, their applications are common when dealing with multivariate continuous variables. Copula estimation with discrete or mixed outcomes is growing but faces challenges of non-uniqueness of copulas that makes interpretation of results difficult. In this chapter, we develop Bayesian copula models for analysing mixed continuous and discrete response in antenatal care (ANC) utilisation. In particular, we constructed three joint models, first to analyse the distribution of mixed binary-continuous data, a second for a mixture of a count and continuous variables, and a third for a discrete set of count and binary variables. The following type of margins, corresponding to the bivariate set of outcomes, were assumed: a Bernoulli and a gamma for the bivariate-continuous outcomes, a gamma and positive Poisson for the count and continuous outcomes and a Bernoulli and Poisson for the binary and count mix. Then, a flexible Matern-family model is added to capture spatial heterogeneity at district level. The models are applied to study ANC utilisation among Malawian women using the 2015 Malawi Demographic and Health Survey (MDHS) data, drawn using a stratified cross-sectional survey design. Results demonstrate that both individual and contextual factors are important in determining factors influencing use of ANC. The results also showed very strong spatial variation in the timing

E. Gondwe (✉)
University of Malawi, Chancellor College, Zomba, Malawi

M. G. Chipeta
AFIDEP, Lilongwe, Malawi

University of Oxford, Big Data Institute, Oxford, UK
e-mail: michael.chipeta@ndm.ox.ac.uk

L. Kazembe
University of Namibia, Windhoek, Namibia
e-mail: lkazembe@unam.na

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_10

of first antenatal visit, frequency of use as well as access to the nearest health facility across the country. Any planning to promote the use of health care must be programmed to enhance both targeted interventions for improved health care and for accelerated achievement of sustainable development goal number 3.

Keywords Bivariate mixture models · Copula-based regression · Antenatal care · Gaussian copula · Frank copula · Student-t copula · Martèrn correlation function · Spatial variation · Malawi DHS · Multivariate distribution · Clayton copula · Generalised linear model · Generalised joint regression model · Smoothing · Gaussian markov random field · Isotropic · Maternal health · Clustering · Spatial dependency · Health care utilisation

1 Background

Health care (HC) utilisation is the measure of the population's use of the health care service available to them. Good utilisation of health care services serves to improve the health status of the population. According to World Health Organization (WHO) report, Malawi health system ranks number 185 out of 190 of the world's health systems (WHO, 2016). Malawi's Ministry of Health is responsible for health care in Malawi and out of the health services that are provided in Malawi, 62 % of them are provided by the government, 37 % are provided by the Christian Health Association of Malawi (CHAM), and a small fraction of the population receive health services through the private Sector (Makwero, 2018). Health care services must be available to all to ensure that healthy lives and promote well-being for all at all ages, which is a sustainable development goal (SDG) number 3, must be acceptable and of high quality in addition to being an accessible distance or travel time and cost. In most developing countries like Malawi, where the burden of ill health is greatest, adequate information on the location of populations, health services facilities must always be there to give a picture on how people utilise health.

Antenatal care (ANC) is one of the maternal health care services provided by skilled health care professionals to pregnant women and adolescent girls in order to ensure the best health conditions for both mother and baby during pregnancy (WHO, 2010). WHO recommends that every pregnant woman should have at least four antenatal visits during pregnancy and should seek ANC within the first 3 months of pregnancy. The health care that a mother receives during pregnancy is important for the survival and well-being of both the mother and child (Amentie et al., 2015). Despite progress being made in many countries in increasing the availability of maternal health care, more women across Africa do not fully utilise the care (Panel, 2010). WHO estimates that more than half a million women lose their lives in the process of reproduction worldwide every year, and most of these mortalities can be avoided if mothers have access to maternal health care services (Zelalem Ayele et al., 2014). For every maternal death, an estimated 30 to 50 women suffer pregnancy-related healthy problem such as vesicovaginal fistulate, infertility

and depression, which could have been avoided if the required ANC visits were achieved (Amentie et al., 2015).

Timing of ANC visits is an appropriate time for one to be aware of signs and symptoms of pregnancy complications, which in turn leads to timely access to appropriate care (Gidey et al., 2017). Early antenatal care attendance during the first 3 months of gestation plays a major role in detecting and treating complications of pregnancy and forms a good basis for appropriate management during delivery and after childbirth. Initiating ANC early may help to prevent stillbirths in term pregnancies (baby born within 37–42 weeks of pregnancy) by preventing labour complications through early referral to skilled birth attendants (Yakoob et al., 2010). Receiving ANC during pregnancy has a positive effect on the utilisation of postnatal health care services, which reduces the odds of neonatal deaths. Postnatal health care is a means of providing follow-up care to newborns and provides an opportunity to check all newborns and the mother for illnesses that may have arisen from the time of birth for the first 6 weeks or so (WHO, 2016).

Understanding relationships among multiple health outcomes is fundamental to improving utilisation and access by the population. The Malawi Health Strategic Plan (2017–2022) highlights the need to promote utilisation of HC services. Particularly, antenatal care is one of the most effective health intervention for preventing maternal morbidity and mortality especially in areas where the general health status of a woman is poor. Most maternal mortalities and disabilities are due to direct complication, which are avoidable if women can get adequate and timely antenatal delivery and post delivery services (Zelalem Ayele et al., 2014).

Joint analysis of ANC visits and timing at first visit is of importance as the ANC visits alone cannot meet the required number of visits recommended by WHO, and timing at first visit can have a contribution to the number of visits a woman can have during the pregnancy period. This means that frequency of ANC visits would largely depend on when the woman started the first visit and late timing may lead to less number of ANC visits, hence allowing to establish the association between these outcomes. Copula models permit a flexible approach to modelling dependence in interrelated outcome variables. Studying patterns of HC utilisation may assist in identifying where resources need to be targeted. Copula-based approaches have proven to be particularly suitable for modelling data showing departures from multivariate normality. Copulas allow us to model separately the marginals from the dependence structure and the use of different copula families. The literature of copula applications is vast. For environmental, actuarial and financial applications, see, for example, Christian Genest (2007), Patton (2006), Jondeau and Rockinger (2006), Umberto Cherubin (2004), among others. A detailed overview and their properties is given by Joe (1997) and Nelsen (2007)

We propose a bivariate mixed outcomes spatial model, which extends the approach introduced by Kazianka and Pilz (2010) and Kazembe (2019). We demonstrate the usefulness of our approach to ANC utilisation, focussing on timing of ANC and number of ANC visits. Particularly, we considered three types of margins: first, a binary-continuous pair, a second a mixture of a count and continuous variables and a third for a discrete set of count and binary variables,

which departs from the gamma and truncated Poisson marginals approach applied by Kazembe (2019).

The rest of the chapter is organised as follows: Sect. 2 introduces the bivariate copula spatial model, Sect. 3 discusses copula-based spatial methods, Sect. 4 is devoted to the ANC application and Sect. 5 is the concluding discussion.

2 Copula-Based Methods

2.1 General Copula Theory

Copulas are functions that separate the marginal distributions from the dependency structure of a given multivariate distribution (Sklar, 1996). Copulas are basically the multivariate distribution function of uniformly distributed random variables on the interval [0, 1]. Copula modelling became widely adopted in the twenty-first century, applied in many fields, but most famously used in finance and insurance to model default correlations (Embrechts, 2009; Chen et al., 2015). They are increasingly being applied in health applications, including health care utilisation (Quinn, 2007).

Let $F(Y_1, Y_2)$ be joint distribution function of random variables Y_1 and Y_2 whose distributions $F_1(Y_1) = P(Y_1 \leq y_1)$ and $F_2(Y_2) = P(Y_2 \leq y_2)$ are either continuous, discrete or a combination. Sklar's theorem states that there exists a unique **Copula function** $\mathscr{C}(\cdot, \cdot)$ such that

$$C(Y_1, Y_2) = F(F_1^{-1}(Y_1), F_2^{-1}(Y_2)), Y_1, Y_2 \in [0, 1]. \tag{1}$$

As an extension to spatial context, let $F_z, z \in D \subset \mathbb{R}^2$ denote the distribution of the random isotropic and stationary process (i.e. constant mean at each location x_i and covariance is only a function of distance between locations x_i) (Bárdossy, 2006). The relation between two locations separated by the vector h is characterised by the bivariate distribution:

$$P(Z(x) \leq z_1, Z(x + h) \leq z_2) = C_h(F_Z(z_1), F_Z(z_2)), \tag{2}$$

whose dependence structure is described by the copula C_h. In this case, the copula becomes a function of the separating vector h (or the separating distance $h := ||h||$ if the random field is isotropic) and does not depend on the location x.

There are several copula functions that can be used in modelling. Radice et al. (2016) and Hasebe (2013) presented some of the common copula functions which are used in this study. The copula functions are summarised in Table 1. Gaussian copula and Student-t copula are examples of implicit copulas, which do not have a closed form and are mostly used in multivariate distribution function (Aas, 2004). In Table 1, θ is the dependence parameter. The Student-t dependence introduces an additional parameter compared to the rest of the copulas which is called the degree

Table 1 Summary of selected copula functions

Copula type	Function $c(u, v)$	θ—domain
Student t-copula	$\int_{-\infty}^{t_{\theta_1}^{-1}(u)} \cdots \int_{-\infty}^{t_{\theta_1}^{-1}(v)} \frac{1}{2\pi(1-\theta_2^2)^{\frac{1}{2}}} \left(1 + \frac{s^2 - 2\theta_2 st + t^2}{\theta_1(1-\theta_2^2)}\right)^{\frac{(\theta_1+2)}{2}} dsdt$	$\theta_1, \theta_2 \epsilon [-1, 1]$
Frank	$-\theta^{-1} log\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{\theta} - 1}\}$	$\theta \epsilon (-\infty, \infty)$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{\frac{-1}{\theta}}$	$\theta \epsilon (0, \infty)$
Gaussian	$\Phi_2(\Phi^{-1}(u), (\Phi^{-1}(v); \theta)$	$\theta \epsilon [-1, 1]$

of freedom. This means that for a student t copula, θ_1 is the parameter of the copula and θ_2 is the degrees of freedom. The Student-t copula allows for joint extreme events but not for asymmetries and may be too restrictive to provide reasonable fit for asymmetries. In this case, a clayton copula that is asymmetric might be a better choice. Clayton copula is one of the explicit copulas which are not derived from a multivariate distribution function but do not have simple closed form (Radice et al., 2016; Hasebe, 2013). Perfect dependence from a clayton copula is obtained if $\theta \to \infty$, while $\theta \to 0$ implies independence. For more of copula theory, refer to (Nelsen, 2007).

2.2 Marginal Regression Models

We develop a bivariate model of three mixed outcomes with marginals following a Bernoulli and Gamma Generalised Linear Models (GLMs), a Gamma and Poisson GLMs and lastly a Bernoulli and Poisson GLMs. The Bernoulli GLM represents whether the four focal number of ANC visits occurred or not (Y_{i1}). The Gamma GLM represents the average time it takes for a pregnant woman to visit ANC (Y_{i2}), while the Poisson GLM represents the number of ANC visits (Y_{i3}).

Suppose (Y_{i1}), (Y_{i2}) and (Y_{i3}) are independent random variables for women $i = 1, \ldots, n$, and then the marginal GLM's are given as

$Y_{i1} \sim Bernoulli(\mu_{i1})$ with $ln(\frac{\mu_{i1}}{1-\mu_{i1}}) = \mathbf{X}_i'\vartheta$
$Y_{i2} \sim Gamma(\mu_{i2}, \nu^2)$ with $ln(\mu_{i2}) = \mathbf{Z}_i'\alpha$
$Y_{i3} \sim Pois(\mu_{i3})$ with $ln(\mu_{i3}) = ln(E_i) + \mathbf{H}_i'\beta$

where $\mathbf{X}$, $\mathbf{H}$ and $\mathbf{Z}$ are distinct or similar set of covariates, with ϑ, α and β corresponding regression associated with discrete Y_{i1}, Y_{i3} and continuous Y_{i2}, respectively, modelled on the means $\mu_{i1} = E[Y_{i1}]$, $\mu_{i2} = E[Y_{i2}]$ and $\mu_{i3} = E[Y_{i3}]$. To account for the expected length of time a visit will occur among a group of pregnant women, we included the offset $ln(E_i)$ in a Poisson model.

The number of antenatal visits categorised into two, i.e. less than 4 visits and at least 4 visits, has Bernoulli density given by

$$f(y_1) = \mu_{i1}^{y_{i1}}(1-\mu_{i1})^{1-y_{i1}} \tag{3}$$

where $y_1 = 0$ or 1.

The average length of time Y_{i2} has a density

$$f_{y2}(y_{i2}|\mu_{i2}, \nu^2) = \frac{1}{\Gamma(\frac{1}{\nu^2})}(\frac{1}{\mu_{i2}\nu^2})^{\frac{1}{\nu^2}} y_{i2}^{\frac{1}{\nu^2}-1} exp\{-\frac{1}{\mu_{i2}\nu^2}y_{i2}\} \tag{4}$$

such that $\text{Var}[Y_{i2}] = \mu_{i2}, \nu^2$ for some parameter ν, which is estimated in the marginal gamma GLM through $\frac{E[Y_{i2}]}{\sqrt{Var[Y_{i2}]}}$. It has to be noted that a visit has to occur first, and then cumulative visits follow. As such the total number of visits considered in this model is only positive counts, hence a density of Poisson given by

$$f_{y3} = \frac{\mu_{i3}^{y_{i3}}}{y_{i3}!(1-e^{-\mu_{i3}})}e^{-\mu_{i3}} \tag{5}$$

for $y_{i3} = 1, 2, \ldots$

3 Copula-Based Spatial Methods

Our focus is on analysing spatial area-referenced data, where $Y_1, \ldots, Y_n$ constitute observed summaries in each area, $i = 1, 2, \ldots, n$. Conditional autoregressive (CAR) models are mostly used to analyse areal-aggregated spatial data. Furthermore, regression models are commonly fitted with one response variable and a set of covariates (Filippou et al., 2019). However, multiple response variables can also be fitted. Joint regression models can be applied to continuous, discrete and mixed correlated outcomes. The most common choice for bivariate outcomes is the Gaussian copula given by

$$C(u_1, u_2|\rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)). \tag{6}$$

The dependency in Eq. (6) has been captured using a bivariate Gaussian copula cumulative distribution function $C(u_1, u_2|\rho)$, where u_1 and u_2 are random variables, Φ is the standard normal distribution and Φ_2 is the standard bivariate normal distribution, with dependence parameter ρ. Song et al. (2009) pointed out that Gaussian copula is an exception because of its flexible dependence structure. Copula dependence and marginal parameters can easily be estimated through the Gaussian method. Each parameter depends on the joint model predictor including different types of covariate effects such as linear, non-linear, random and spatial

effects (Marra and Radice, 2017). A joint cumulative distribution function (CDF) of two response variables Y_1 and Y_2 can be expressed as

$$F(y_1, y_2) = C(F_1(y_1 u_1, \sigma_1, v_1), F_2(y_2 u_2, \sigma_2, v_2); \xi, \theta) \tag{7}$$

where

$(u_1, \sigma_1, v_1, u_2, \sigma_2, v_2, \xi, \theta)^T$, $F_1(y_1|u_1, \sigma_1, v_1)$ and $F_2(y_2|u_2, \sigma_2, v_2)$

are marginal CDFs of Y_1, Y_2 and u_m, σ_m and v_m for $m = 1, 2$ are marginal distribution parameters, C is a defined 2-place copula function with dependence coefficient θ and ξ is the number of degree of freedom (Marra and Radice, 2017).

The correlation function of the Gaussian spatial process could be used to specify well-structured spatial dependency models. Equally important, the use of Gaussian process is compatible with the rich literature on spatial modelling and incorporates the popular basic spatial random effects model as a special case (Banerjee et al., 2008). The Gaussian copula process is specified as follows: let $U_i = \Phi(Z(s_i))$, where $\Phi()$ is the cumulative distribution function (CDF) of $N(0, 1)$. We define

$$Z(s_i) = W(s_i) + \epsilon(s_i), W(s) \sim GP(0, \alpha\rho(s, s'; (v, \varphi))), \epsilon(s_i) \underrightarrow{iid} N(0, 1 - \alpha) \tag{8}$$

where $\rho(, ; (v, \varphi))$ can be any valid correlation function with smoothness parameter v and decay parameter φ. The process W(s) captures structural spatial association, while $\epsilon(s)$ is uncorrelated pure error. The parameter $\alpha \ \epsilon[0, 1]$ determines the proportion of variation that is spatially structured. We adopt the Matérn correlation function to calculate correlations from distances. The Matérn function describes realisations of Gaussian spatial processes with smoothnesses (Pulang and Bhadra, 2022; Guttorp and Gneiting, 2006). In terms of joint modelling, a maximum likelihood approach is used as it can combine random effects.

4 Application to Antenatal Care Data

4.1 Methodology

The study used secondary data, which was collected from the 2015–2016 Malawi Demographic Health Surveys (MDHS), which included Global Positioning System (GPS) coordinates information by the Malawi National Statistical Office and the Malawi Ministry of Health. The cross-sectional survey data were collected between the 19 October 2015 and 17 February 2016. Maternal health data for 9228 reproductive women from 15 to 49 years were used to jointly model the number of antenatal care (ANC) visits a woman has and the timing at her first antenatal visit in the 9-month period of pregnancy. All women residing in the same cluster have the same geo-reference location. To protect the confidentiality of DHS

survey respondents, the geo-located data are displaced before being provided to the researchers such that urban clusters are displaced up to 2 km and rural clusters up to 5 km (Burgert et al., 2013; Warren et al., 2016). Permission to use the data was granted by the Measure DHS programme. Data include demographics, socio-economic determinants and location data. Analysis was done in open-source statistical environment **R** 3.6.1 (R Core Team, 2019)

4.2 Bivariate Copula Spatial Modelling

We jointly model two health outcomes; frequency of antenatal care visits and timing at the first antenatal visit. Three joint models are constructed, first to analyse the distribution of mixed binary-continuous data, a second for a mixture of a count and continuous variables and a third for a discrete set of count and binary variables. The following type of margins, corresponding to the bivariate set of outcomes, are assumed: first we assume a Bernoulli for achieving focused antenatal care or not and a gamma for timing of first antenatal visit giving the bivariate discrete-continuous outcomes, second set of marginals assumed are a gamma for frequency of visits and positive Poisson for the counts, which presents a and continuous-discrete paired set of outcomes, and third assumption is to have a Bernoulli and Poisson marginals for the binary and count mix. To perform a flexible spatial modelling at district level, copula was applied to account for spatial dependency and spatial distribution.

The study used Generalised Joint Regression Model and the function $gjrm$ in **R** statistical software environment. Generalised Joint Regression Model (GJRM) implements a flexible joint modelling framework for fitting a number of multivariate response regression models under various sampling schemes (Marra and Radice, 2017). The framework allows both Gaussian and non-Gaussian dependencies through the use of copula. The copula model used areal data to account for spatial dependence and spatial regression inference. Areal data are known to be neighbours as they share a common boundary making it necessary for joint analysis (Bivand et al., 2013). To account for geographic clustering of antenatal care, a Markov random field approach was employed to exploit the information contained in neighbouring observations which are located in the same district.

Variable $district$ was fitted as an unstructured random effect in the model. A Gaussian Markov random field assumes that a random variable associated with a certain region depends primarily on its neighbours.

Let the neighbours N_i to a point s_i be the points s_j, $j \in N_i$ that are "close" to s_i. A Gaussian random field $x \sim N(\mu, \Sigma)$ that satisfies

$$p(x_i|x_j \neq i) = p(x_i|x_j : j \in N_i) \tag{9}$$

is a Gaussian Markov random field.

Smoothing was also applied to the model for complexity. The smoothing penalty is based on the neighbourhood structure of the geographic units, so that spatially adjacent regions share similar effects. Districts were applied as smooth function to model spatial effects.

The equation in general is defined as

$$y_{xi} = \mathbf{v}_{xi}^T \boldsymbol{\gamma}_x + \Sigma_{k_x=1}^{K_x} s_{xk_x}(\mathbf{z}_x k_{xi}), i = 1, \ldots, n, \forall_x = 1, 2 \tag{10}$$

where n is the sample size, y_{xi} is a latent continuous variable, $\mathbf{v}_{xi}^T$ contains binary and/or categorical predictors, γ_x is the vector that represent the effect of the variable in v_{xi}, $z_x k_{xi}$ denotes the k_x^{th} sub vector of the complete covariate vector z_{xi} containing binary, categorical and spatial variables and K functions and $s_{xk_x}(z_{xk_x i})$ represents the generic effects of the smoothing function depending on the type of covariates that are considered. The latent variable determines the outcome variables (Marra and Radice, 2017). In the equation $x = 1, 2$ is the number of outcome variables, and in this case, 1 is for frequency of ANC visits and 2 is for timing at first visit. Equation (10) is then further extended to include

```
s(x, bs="mrf", xt=xt, k=7)
```

where x is a factor variable, in this case districts, k is the number of folds and mrf stands for Markov random field. The neighbourhood structure information is stored in an object xt, which is then used in specification of the Gaussian Markov random field smoother. Based on the 2015 MDHS data, there are 28 geographic districts in Malawi. This allowed us to use the information contained in the neighbouring observation that is located in the same country as spatial pattern suggests that areal observations close to each other are more similar than those that are far from each other. The two outcomes in this chapter are the frequency of antenatal visit and timing at first visit whose margins are Bernoulli and gamma, respectively, which includes spatial effects.

4.3 Results

Description of Key Variables

Variables in the analysis include age of a woman, region, place of residence, sex of household head, frequency of listening to radio, frequency of reading newspaper, if the woman is currently working, timing at first visit, household wealth status, the number of antenatal care visits, education level, if a woman ever terminated pregnancy before and whether she wanted the pregnancy. The two outcomes of interest are timing at first visit and the number of antenatal visits, i.e. whether

Table 2 Characteristics of the study participants

Variable	Visit Frequency		
	< 4 visits [N(%)]	≥ 4 visits [N(%)]	Total [N(%)]
All	4361	4867	9228
Age			
15–21	803 (18)	757 (16)	1560 (17)
22–28	1618 (37)	1711 (35)	2329 (36)
29–35	1206 (28)	1516 (31)	2722 (30)
36–42	600 (14)	698 (14)	1298 (14)
43–49	134 (3)	185 (4)	319 (3)
Region			
Central	1610 (37)	1881 (37)	3491 (38)
Northern	775 (18)	867 (18)	1642 (18)
Southern	1976 (45)	2121 (44)	4095 (44)
Residence			
Rural	3801 (87)	3999 (82)	7800 (85)
Urban	560 (13)	868 (17)	1428 (15)
Household head sex			
Female	1141 (26)	1303 (27)	2444 (26)
Male	3220 (74)	3564 (73)	6784 (74)
Radio listening			
≥ once a week	1271 (29)	1639 (34)	2910 (32)
< once a week	767 (18)	893 (18)	1660 (18)
Not at all	2323 (53)	2335 (48)	4658 (50)
Reading newspaper			
≥ once a week	1271 (29)	1639 (34)	2910 (32)
< once a week	767 (18)	893 (18)	1660 (18)
Not at all	2323(53)	2335 (48)	4658 (50)
Education			
No education	570 (13)	527 (10)	1097 (12)
Primary	2921 (67)	3095 (64)	6016 (65)
Secondary	822 (19)	1085 (22)	1907 (21)
Higher	48 (1)	160 (3)	208 (2)
Wealth Index			
Poor	1054 (24)	1007 (21)	2061 (22)
Poorer	969 (22)	974 (20)	1943 (21)
Middle	892 (21)	952 (19)	1844 (20)
Richer	825 (19)	907 (19)	1732 (19)
Richest	621 (14)	1027 (21)	1648 (18)
Terminated pregnancy			
Yes	452 (10)	548 (11)	1000 (11)
No	3909 (90)	4319 (89)	8228(89)

(continued)

Table 2 (continued)

Variable	Visit Frequency		
	< 4 visits [N(%)]	≥ 4 visits [N(%)]	Total [N(%)]
Wanted pregnancy			
Later	1448 (33)	1369 (28)	2817 (31)
No more	582 (13)	606 (12)	1188 (13)
Then	2331 (53)	2892 (59)	5223 (57)
Timing at first visit			
1st trimester	503 (12)	1903 (39)	2406 (26)
2nd trimester	3335 (70)	2923 (60)	6258 (68)
3rd trimester	523 (12)	41 (1)	564 (6)
Currently working			
Yes	4066 (93)	4578 (94)	8644 (94)
No	295 (7)	290 (6)	585 (6)

one has less than 4 visits or more than or equal to 4 visits. Table 2 shows the description of the variables that were used in the study. The choice of the covariates was based on the studies by Machira (2017); Oladipo (2014) as they were shown to be significant determinants of maternal health care utilisation. Out of the 9228 women, 47% of them had less than four visits and 53% had more than or equal to four 4 visits. Of the total, 85% were from rural areas and 15 % were from urban.

Modelling

Different copula models were fit to identify one that best fits the data at hand. Four of the commonly known copula models were identified which include Gaussian, Clayton, Frank and Student-t. Table 3 shows performance results from the selected copulas together with their akaike information criterion and the dependence measure θ. In the table, θ is the dependence measure, δ^2 is the variance and τ is the kendal's tau value. From the table below, the preferred copula model is the Student-t since it has the lowest Akaike Information Criterion (AIC) of 41191.38. The Student-t copula showed a dependence measure of -0.58, and 95% confidence interval ranges from -0.60 to -0.563. The best copula was therefore fitted to the data for further analysis.

Table 4 shows a summary of Student-t performance on the three different marginal combination. The estimated dependence measure in all the three models showed a negative dependence between frequency of antenatal care visit and timing at first visit. Based on the AIC results, the best fitting model to the data (joint outcomes) is the Bernoulli–Gamma marginal combination, with the lowest AIC of 41028.

Table 3 Copula functions performance

Model	df	AIC	δ^2 (95% CI)	θ (95% CI)	τ (95% CI)
Gaussian	50	41290.6	0.087	−0.544	−0.0367
			(0.084, 0.089)	(−0.57, −0.53)	(−0.383, −0.354)
Clayton	50	43253.24	0.086	0.001	0.001
			(0.084, 0.089)	(0.001, 100)	0.001, 0.98)
Frank	50	41224.85	0.087	−3.6	−0.358
			(0.085, 0.089)	(−3.76, −3.46)	(−0.37, −0.346)
Student-*t*	50	41191.38	0.087	−0.58	−0.39
			(0.085, 0.089)	(−0.60, −0.563)	(−0.411, −0.38)

List of abbreviations
df = Degree of freedom
AIC = Akaike Information Criterion
CI = Confidence Interval

Table 4 Summary of Student-*t* copula performance on three different marginal combination

Model	df	AIC	θ (95% CI)
Bernoulli–Gamma	57	41,028	−0.580(−0.596, −0.556)
Bernoulli–Poisson	58	44,633	−0.745(−0.759, −0.731)
Poisson–Gamma	57	60,645	−0.707(−0.717, −0.696)

List of abbreviations
df = Degree of freedom
AIC = Akaike Information Criterion
CI = Confidence Interval

4.4 Bivariate Model Results

By the time MDHS 2015–2016 was conducted, Malawi was still following WHO recommendation that every pregnant woman should have at least 4 antenatal care visits during pregnancy period. In line with this, we recode frequency of antenatal care visit as binary variable, indicating whether one has less than 4 or at least 4 visits as recommended by the world Health Organization (WHO) during the time these data were collected. The joint outcomes were modelled and copula estimates were obtained.

The results in Table 5 show that the covariates which were seen to be significant with ANC visits (those that have their p-values less than 0.05) were region, educational level, age of a woman, wealth index, if a woman ever terminated pregnancy, place of residence and if the woman wanted the pregnancy. Similarly, covariates that were seen to be significant with timing at first visit (those that have their p-values less than 0.05) were region, educational level, age of a woman, whether a woman was working or not and wealth index. The Bernoulli–Gamma model showed a dependence measure of −0.58 with a 95 % confidence interval

Table 5 Spatial copula model estimates for Bernoulli–Gamma margins

	ANC visits			Timing at first visit		
Parameter	Estimate	Std error	p-value	Estimate	Std error	p-value
Intercept	−0.020	0.114	0.858	1.567	0.0278	< 0.001
Region: Ref.=*North*						
Central	−0.301	0.092	0.001	−0.042	0.021	0.045
South	−0.234	0.118	0.048	−0.046	0.027	0.093
Education: Ref.=*None*						
Primary	0.099	0.044	0.023	−0.012	0.010	0.217
Secondary	0.162	0.054	0.002	−0.030	0.012	0.015
Tertiary	0.577	0.111	< 0.001	−0.156	0.024	< 0.001
Age (years):Ref.=*15 - 21*						
22–28	0.014	0.038	0.718	0.0112	0.010	0.244
29–35	0.131	0.041	0.001	0.019	0.011	0.041
36–42	0.130	0.050	0.010	0.035	0.012	0.003
43–49	0.329	0.082	< 0.001	0.025	0.019	0.180
Working: Ref.=*No*						
Yes	0.084	0.054	0.126	−0.035	0.013	0.010
HH: Ref.=*Male*						
Female	0.052	0.031	0.095	0.003	0.007	0.716
Wealth: Ref.=*Poorer*						
Poor	0.040	0.040	0.316	−0.015	0.009	0.107
Middle	0.067	0.0413	0.1240	−0.023	0.010	0.019
Richer	0.052	0.041	0.243	−0.024	0.010	0.022
Richest	0.198	0.055	< 0.001	−0.055	0.013	< 0.001
Residence: Ref.=*Rural*						
Urban	0.087	0.044	0.047	0.010	0.010	0.312
Radio: Ref.=*No*						
< a week	0.067	0.037	0.072	−0.001	0.01	0.863
≥ once a week	0.062	0.032	0.055	−0.015	0.008	0.052
Newspaper: Ref.=*No*						
< a week	0.001	0.044	0.984	0.009	0.010	0.355
≥ once a week	0.029	0.055	0.589	−0.024	0.013	0.062
Evertempreg: Ref.=*No*						
Yes	0.035	0.042	0.433	−0.008	0.009	0.367
Wantedpreg: Ref.=*Then*						
No more	−0.159	0.044	< 0.001	0.044	0.010	< 0.001
Later	−0.145	0.029	< 0.001	0.045	0.007	< 0.001

List of abbreviations
HH = Household head
Radio = Listening to radio
Newspaper = Reading news paper
Evertempreg = Ever terminated pregnancy

of (−0.595, −0.565). The estimated dependence parameter from the joint model supports the hypothesis that those who are most likely to have more number of antenatal visits are those who are likely to have their first visit during the first trimester as it showed a negative dependency.

Mother's education is seen to be significant in both models. The estimates suggest that women with primary education were more likely to have more than or equal to 4 visits and would attend the first ANC visit early in the pregnancy, with an estimate of 0.014 (p-value 0.023) than those with no education. A similar trend was observed with increasing levels of education, secondary education (0.162, p-value 0.002) and tertiary (0.577, p 0.001).

Results indicate that age was associated with frequency of ANC visits and timing at initial visit. Older mothers were more likely to achieve at least 4 ANC visits (0.329, p-value 0.001) compared to reference category of young mothers aged 15–21 years.

Family wealth status was shown to be positively associated with both increased number of ANC visits and early initiation of ANC. Women from richest households were more likely to have at least 4 visits (0.198, p-value 0.001) and early start on ANC (-0.055, p-value 0.001)

Women who no longer wanted the pregnancy were less likely to achieve at least 4 ANC visits (-0.159, p-value 0.001). Similarly, women who wanted to get pregnant later were shown to have low ANC visit frequency (-0.145, p-value 0.001). However, in terms of ANC initiation, women who did not want pregnancy tended to initiate early in time (0.044, p-value 0.001). Similarly, a woman who later wanted the pregnancy has a bit high chances (0.007, p-value 0.001) of initiating the antenatal care visit late than the woman who wanted the pregnancy then.

Estimates show that women from central region and southern regions were less likely to have at least 4 ANC visits by 0.301 (p-value 0.001) and 0.234 (p-value 0.048), respectively, than women from the northern region. Region was seen to be significant in explaining a woman's timing at her first visit, and estimates show that women from central region were more likely to have a first visit late with an estimate of 0.021 than women from the north. Similarly, women from southern region are more likely to start their first visit late with an estimate of 0.027 than those from the north.

Place of residence was seen to be significant with antenatal care visit. Estimates show that women from urban are more likely to have more ANC visits (0.087 p-value 0.047) than those from the rural areas. Similarly, results from Poisson-Gamma and Bernoulli-Poisson combinations shows similar pattern; see Tables 6 and 7 in the appendix.

4.5 Spatial Variation Across Malawi

This subsection presents high-resolution maps for antenatal visit and timing at first visit. Figure 1 shows district spatial variation of ANC visits across the country. The p-value of the smooth function equation less than 0.001, which showed to be significant in explaining the spatial dependence. In Fig. 1, geographical locations show to have contributed to variations in ANC visits frequency across the country. Districts with lighter shade simply means that there is a higher utilisation of ANC compared with the darker shades. The darker the shade, the lower the ANC utilisation. In both Figs. 1 and 2, $s(dist\_code, 5.75)$ simply means the smooth district function that was used in the model. Chitipa, Karonga and Rumphi have the lowest ANC utilisation while Mzimba, Nkhatabay, Nsanje, Chikwawa, Mwanza and Blantyre has a moderate spatial variation. Most districts in the central region show a higher ANC utilisation. However, Phalombe shows the highest ANC utilisation. Most parts of the northern region of Malawi showed to have the lowest ANC utilisation.

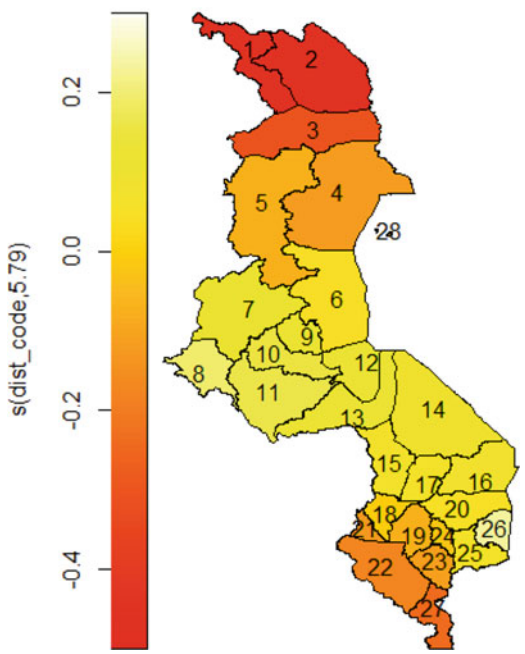


Fig. 1 Spatial variability for ANC visits frequency: Chitipa (1), Karonga (2), Rumphi (3), NkhataBay (4), Mzimba (5), Nkhotakota (6), Kasungu (7), Mchinji (8), Ntchisi (9), Dowa (10), Lilongwe (11), Salima (12), Dedza (13), Mangochi (14), Ntcheu (15), Machinga (16), Balaka (17), Neno (18), Blantyre (19), Zomba (20), Mwanza (21), Chikwawa (22), Thyolo (23), Chiradzulu (24), Mulanje (25), Phalombe (26), Nsanje (27) and Likoma (28)

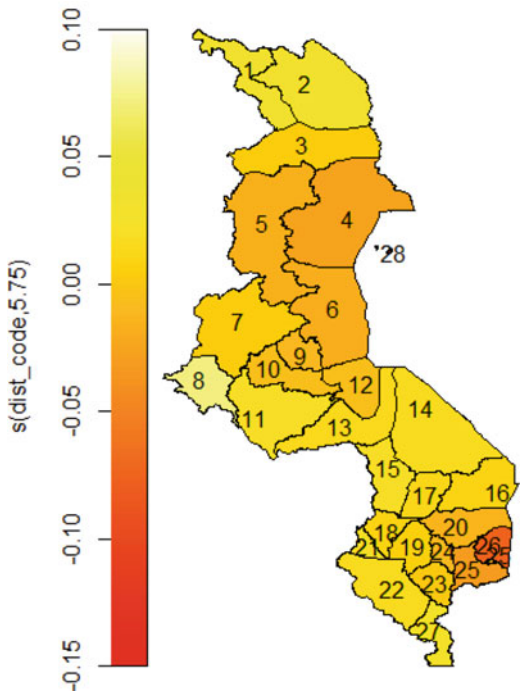


Fig. 2 Spatial variability for timing at first visit; Chitipa (1), Karonga (2), Rumphi (3), NkhataBay (4), Mzimba (5), Nkhotakota (6), Kasungu (7), Mchinji (8), Ntchisi (9), Dowa (10), Lilongwe (11), Salima (12), Dedza (13), Mangochi (14), Ntcheu (15), Machinga (16), Balaka (17), Neno (18) Blantyre (19), Zomba (20), Mwanza (21), Chikwawa (22), Thyolo (23), Chiradzulu (24), Mulanje (25), Phalombe (26), Nsanje (27) and Likoma (28)

Figure 2 shows district spatial variation of timing at first visit. The p-value of the smooth function equations was less than 0.001, which showed to be significant in explaining the spatial dependence. Geographical locations showed to have contributed to variations in timing at first visit across the country. Women from Phalombe show to have started their first visit late, and Nkhatabay, Mzimba, Nkhotakota, Mulanje, Ntchisi, Dowa and Salima women show to have started their first ANC visit a bit early. Women from Mchinji show to have started their first visit very early followed by Karonga, Chitipa, Nsanje and Chikwawa.

Discussion

The results from the joint model shows that region, educational level, age of a woman, wealth index and sex of household head were significant with antenatal care visits and timing at first visit. In addition to that, if a woman ever terminated pregnancy, exposure to radio, place of residence and if the woman wanted the pregnancy were also significant with antenatal care visits and timing at first visit. These results are similar to studies done by Tekelab et al. (2019) and Gupta et al. (2014). Both studies showed that education level, planned or wanted pregnancy were associated with the number of antenatal visits. Education is likely to increase individual's awareness of health care services and the benefits of initiating antenatal care early in turn having the required number of visits. Educated women usually

have a greater awareness of the existence of ANC services and the advantages of using such services (Efendi et al., 2017). They are more aware of health problems, know more about the availability of health care services and utilize the information more effectively than non-educated women (Zhao et al., 2012).

According to Roberts et al. (2017), mother's education was the most consistent and important determinant of the use of child and maternal health services which is also similar to the study by Ajayi and Osakinle (2013), which showed that respondent with no education was 1.9 times less likely to have the required number of visits. Those with the highest level of education were 1.3 times more likely to have an increase in the number of ANC visits.

In the joint model, estimates showed that the place of residence is significant with ANC visits. This concurs with a study by Edward (2011), which looked at factors influencing the utilisation of antenatal care in Uganda. The study showed that location differences are revealed to be significant in influencing the utilisation of antenatal care content. Being in the rural area, compared to one in the urban area, reduced the utilisation of antenatal care by 0.3 ($p < 0.01$) to 0.4 ($p < 0.01$) percentage points.

Despite the introduction of free maternal health care in recent times, results show that the level of ANC attendance increases with increase in wealth. This finding is similar to a study done by Gebremeskel et al. (2015), which showed that women with low monthly income had higher odds of late antenatal care attendance, hence less number of visits. This is also consistent with a study conducted in Laos by Ye et al. (2010), which found that women with higher income are more likely to utilise antenatal services than those with less income. This may be the case since rich or wealthy people have all the resources to be used for instance, if the hospital where the woman attends her ANC is far from her home, to the rich it is easier to find their way to the hospital since the transport would always be there. Unlike the poor if they do not have any means of transport that might hinder the attendance of ANC and may end up having a less number of visits due to lack of transport.

High-resolution spatial maps showing ANC utilisation reveal that there is spatial variation across the districts in the country. Most part of the northern region of Malawi showed to have the lowest utilisation of antenatal care. The central part of Malawi is a bit flat with well-connected roads, and the southern part has a lot of mountains which could also lead to the variation in accessing antenatal care. The spatial variation of ANC visits and timing at first visit could also be as a result of having health facilities far from the place of residence due to its geographical setup. Distance to the nearest health facility has proven to be an obstacle in utilising antenatal care services. Ali et al. (2018) revealed a very strong association between distance and attendance of ANC. In most cases, distance has been identified as an important barrier to the use of health services, especially in rural areas (Noorali et al., 1999). Kambala et al. (2011) also identified long distances to access health centres as one factor that hinders pregnant women to use antenatal clinics (ANC), delivery and postnatal care in Chikhwawa district in Malawi. Kim et al. (2019)

explored the role of health facility availability as it relates to maternal health care use in rural Malawi. In terms of health facility availability, nearly 32 % of rural women had no health facilities within 5 km of where they lived, with the central and northern regions having significantly higher percentages with no facilities (37.2 and 32.5 %, respectively). This agrees with our findings that showed that Chitipa, Karonga and Rumphi have the lowest spatial variation of ANC visits, while Mzimba and Nkhatabay had a moderate spatial variation.

5 Conclusion and Recommendation

We described the bivariate copula spatial models implemented in the R add-on package GJRM. The framework allowed us to specify flexibly covariate effects and the dependence structure between the specified margins. In this study, we explored the copula dependence between Bernoulli and gamma margins for the bivariate-continuous outcomes, a gamma and positive Poisson for the count and continuous outcomes and a Bernoulli and Poisson for the binary and count. In the Bernoulli–Gamma model, the dependence measure showed a negative association which indicates that those who had four or more antenatal care visits as recommended by WHO were likely to have started their first visit early (first trimester) than those who had less than 4 visits.

Despite the introduction of free ANC services by the Government of Malawi in 2004, the level of accessing ANC equally is still low in the country. The results of this study demonstrate that both individual and contextual factors are important in determining factors influencing use of ANC. Results have also shown that timing of first visits across Malawi to be crucial in the utilisation of ANC services. Any planning to promote use of health care must be programmed to enhance both, targeted interventions for improved health care and for accelerated achievement of sustainable development goals. The study recommends that strategies should be developed for empowering communities to overcome obstacles to reach ANC. These may include using community channels to identify pregnant women, targeting those more likely to be non-users, such as adolescents and women who are poor and single, and making the services more responsive to the needs of women. In addition, introduction of mobile ANC services so as to ease travelling hiccups, especially for women in hard-to-reach and rural areas without health facilities. Further research should be conducted to include survey weights in the analysis. In addition to that, further research should be conducted by using other joint analysis models other than copula and one can also compare the performance of copula models and classic joint models.

Appendices

(Tables 6 and 7)

Table 6 Spatial copula model estimates for Poisson-gamma margins

	ANC visits (Poisson)			Timing at first visit (gamma)		
Parameter	Estimate	Std error	p-value	Estimate	Std error	p-value
Intercept	1.286	0.044	< 0.001	1.592	0.031	< 0.001
Region: Ref.=*North*						
Central	−0.057	0.034	0.089	−0.039	0.024	0.097
South	−0.034	0.045	0.455	−0.043	0.031	0.169
Education: Ref.=*None*						
Primary	0.029	0.017	0.106	−0.011	0.012	0.344
Secondary	0.045	0.021	0.041	−0.029	0.015	0.053
Tertiary	0.149	0.039	< 0.001	−0.145	0.028	< 0.001
Age(years):Ref.=*15 - 21*						
22–28	−0.009	0.016	0.552	0.012	0.010	0.244
29–35	0.019	0.017	0.252	0.227	0.113	0.045
36–42	0.019	0.020	0.335	0.037	0.014	0.007
43–49	0.071	0.033	0.002	0.029	0.022	0.191
Working: Ref.=*No*						
Yes	0.012	0.022	0.578	−0.034	0.015	0.021
HH: Ref.=*Male*						
Female	0.014	0.013	0.271	0.001	0.008	0.921
Wealth: Ref.=*Poor*						
Poorer	0.011	0.017	0.494	−0.016	0.011	0.015
Middle	0.018	0.017	0.288	−0.023	0.012	0.045
Richer	0.017	0.018	0.357	−0.023	0.012	0.062
Richest	0.073	0.022	< 0.001	−0.059	0.015	< 0.001
Residence: Ref.=*Rural*						
Urban	0.013	0.017	0.444	0.012	0.011	0.311
Radio: Ref.=*No*						
$<$ a week	0.009	0.015	0.554	0.002	0.010	0.852
$\geq$ once a week	0.009	0.013	0.495	−0.012	0.009	0.168
Newspaper: Ref.=*No*						
$<$ a week	0.001	0.044	0.984	0.009	0.010	0.355
$\geq$ once a week	0.029	0.055	0.589	−0.024	0.013	0.062
Evertempreg: Ref.=*No*						
Yes	0.012	0.017	0.479	−0.008	0.012	0.444
Wantedpreg: Ref.=*Then*						
No more	−0.044	0.018	0.014	0.043	0.012	< 0.001
Later	−0.053	0.012	< 0.001	0.044	0.008	< 0.001

Table 7 Spatial copula model estimates for Bernoulli–Poisson margins

	ANC visits (Bernoulli)			Timing at first visit (Poisson)		
Parameter	**Estimate**	**Std error**	p-**value**	**Estimate**	**Std error**	p-**value**
Intercept	0.004	0.101	0.973	1.560	0.026	< 0.001
Region: Ref.=*North*						
Central	−0.223	0.080	0.006	−0.042	0.021	0.043
South	−0.168	0.104	0.108	−0.046	0.027	0.093
Education: Ref.=*None*						
Primary	0.076	0.039	0.052	−0.012	0.010	0.217
Secondary	0.131	0.048	0.001	−0.030	0.012	0.015
Tertiary	0.046	0.098	< 0.001	−0.156	0.024	< 0.001
Age(years):Ref.=*15–21*						
22–28	0.013	0.038	0.728	0.012	0.011	0.244
29–35	0.131	0.041	0.001	0.019	0.011	0.041
36–42	0.130	0.050	0.010	0.035	0.012	0.003
43–49	0.329	0.082	< 0.001	0.025	0.019	0.180
Working: Ref.=*No*						
Yes	0.082	0.054	0.127	−0.035	0.013	0.010
HH: Ref.=*Male*						
Female	0.052	0.031	0.092	0.003	0.007	0.716
Wealth: Ref.=*Poor*						
Poorer	0.040	0.040	0.316	−0.015	0.009	0.107
Middle	0.067	0.041	0.107	−0.023	0.010	0.019
Richer	0.052	0.044	0.243	−0.024	0.010	0.022
Richest	0.198	0.055	< 0.001	−0.055	0.013	< 0.001
Residence: Ref.=*Rural*						
Urban	0.087	0.044	0.047	0.010	0.010	0.313
Radio: Ref.=*No*						
$<$ a week	0.067	0.037	0.072	−0.001	0.009	0.863
$\geq$ once a week	0.062	0.032	0.055	−0.015	0.008	0.052
Newspaper: Ref.=*No*						
$<$ a week	0.001	0.044	0.984	0.009	0.010	0.355
$\geq$ once a week	0.029	0.055	0.589	−0.024	0.013	0.062
Evertempreg: Ref.=*No*						
Yes	0.035	0.042	0.479	−0.008	0.012	0.367
Wantedpreg: Ref.=*Then*						
No more	−0.159	0.043	< 0.001	0.044	0.010	< 0.001
Later	−0.144	0.029	< 0.001	0.045	0.007	< 0.001

References

Aas, K. (2004). *Modelling the dependence structure of financial assets: a survey of four copulas*. Norway: NR Note.

Ajayi, I. O., & Osakinle, D. C. (2013). Factors determining the adequacy of antenatal care among pregnant women visiting Ekari state primary health centres. *Online Journal of Health and Applied Sciences, 12*(2), 1–6.

Ali, S., Dero, A., & Ali, S. (2018). Factors affecting the utilization of antenatal care among pregnant women: A literature review. *J Preg Neonatal Med 2018, 2*(2), 41–45.

Amentie, M., Abera, M., & Abdulahi, M. (2015). Utilization of antenatal care services and influencing factors among women of child bearing age in Assosa, Benishangul Gumuz regional state, West Ethiopia. *The Journal of Medical Research, 15*.

Banerjee, O., Ghaoui, L. E., & D'aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research, 9*(Mar), 485–516.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research, 42*(11), 8920–8937.

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Modelling areal data* (pp. 263–318). Berlin: Springer.

Burgert, C. R., Colston, J., Roy, T., & Zachary, B. (2013). *Geographic displacement procedure and georeferenced data release policy for the demographic and health surveys*. New York: ICF International

Chen, Q.-a., Wang, D., & Pan, M. (2015). Multivariate time-varying copula GARCH model and its application in the financial market risk measurement. *Mathematical Problems in Engineering*, 286014, p. 9.

Christian Genest, A. C. F. (2007). Everything you always wanted to know about copula Modeling but you were afraid to ask. *Journal of Hydrologic Engineering, 12*(4), 347–368.

Edward, B. (2011). Factors influencing the utilization of antenatal content in Uganda. *The Australasian Medical Journal, 4*(9), 516.

Efendi, F., Chen, C.-M., Kurniati, A., & Berlian, S. M. (2017). Determinants of utilization of antenatal care services among adolescent girls and young women in Indonesia. *Women and Health, 57*(5), 614–629.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance, 76*(3), 639–650.

Filippou, P., Kneib, T., Marra, G., & Radice, R. (2019). A trivariate additive regression model with arbitrary link functions and varying correlation matrix. *Journal of Statistical Planning and Inference, 199*, 236–248.

Gebremeskel, F., Dibaba, Y., & Admassu, B. (2015). Timing of first antenatal care attendance and associated factors among pregnant women in Arba Minch town and Arba Minch district, Gamo Gofa zone, South Ethiopia. *Journal of Environmental and Public Health, 2015*, 7.

Gidey, G., Hailu, B., Nigus, K., Hailu, T., & Gerensea, H. (2017). Timing of first focused antenatal care booking and associated factors among pregnant mothers who attend antenatal care in central zone, Tigray, Ethiopia. *BMC Research Notes, 10*(1), 608.

Gupta, S., Yamada, G., Mpembeni, R., Frumence, G., Callaghan-koru, J. A., Stevenson, R., Brandes, N., & Baqui, A. H. (2014). Factors associated with four or more antenatal care visits and its decline among pregnant women in Tanzania between 1999 and 2010. *Plos One, 9*(7), e101893.

Guttorp, P., & Gneiting, T. (2006). *Studies in the history of probability and statistics xlix on the Matern correlation family*.

Hasebe, T. (2013). Copula-based maximum-likelihood estimation of sample-selection models. *The Stata Journal, 13*(3), 547–573.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. New York: Chapman and Hall/CRC.

Jondeau, E., & Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance, 25*(5), 827–853.

Kambala, C., Morse, T., Masangwi, S., & Mitunda, P. (2011). Barriers to maternal health service use in Chikwawa, Southern Malawi. *Malawi Medical Journal, 23*(1), 1–5.

Kazembe, L. N. (2019). *Bivariate copula-based regression to model timing and frequency of antenatal care utilization* (pp. 263–318). New York: Nova Science Publisher.

Kazianka, H., & Pilz, J. (2010). Copula-based geostatistical modelling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment, 24*(5), 661–673.

Kim, E. T., Singh, K., Speizer, I. S., Angeles, G., & Weiss, W. (2019). Availability of health facilities and utilization of maternal and newborn postnatal care in rural Malawi. *BMC Pregnancy and Childbirth, 19*(1), 503.

Machira, K. (2017). Determinants of maternal health care services utilization in Malawi. *Public Health* (pp. 805–820).

Makwero, M. T. (2018). Delivery of primary health care in Malawi. *African Journal of Primary Health Care and Family Medicine, 10*(1), 1–3.

Marra, G., & Radice, R. (2017). A joint regression modelling framework for analyzing bivariate binary data in R. *Dependence Modeling, 5*(1), 268–294.

Nelsen, R. B. (2007). *An introduction to copulas*. Berlin: Springer.

Noorali, R., Luby, S., & Hossein Rahba, M. (1999). Does use of a government service depend on distance from the health facility? *Health Policy and Planning, 14*(2), 191–197.

Oladipo, J. (2014). Utilization of health care services in rural and urban areas; a determinant factor in planning and managing health care delivery system. *Africa Health Sciences, 14*(2), 322–333.

Panel, A. P. (2010). *Maternal health: Investing in the lifeline of healthy societies and economics policy brief*.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review, 47*(2), 527–556.

Pulang, M., & Bhadra, A. (2022). Beyond matèrn: On a class of interpretable confluent hypergeometric covariance functions. *Journal of the American Statistical Association*, 1–14. https://doi.org/10.1080/01621459.2022.2027775

Quinn, C. (2007). *The Health Economic Applications of Copulas: Methods in Applied Econometrics Research*. PhD thesis.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radice, R., Marra, G., & Wojtyś, M. (2016). Copula regression spline models for binary outcomes. *Statistics and Computing, 26*(5), 981–995.

Roberts, J., Hopp Marshak, H., Sealy, D.-A., Manda-Taylor, L., Mataya, R., & Gleason, P. (2017). The role of cultural beliefs in accessing antenatal care in Malawi: a qualitative study. *Public Health Nursing, 34*(1), 42–49.

Sklar, A. (1996). Random variables, distribution functions, and copulas: A personal look backward and forward. *Lecture Notes-Monograph Series, 28*, 1–14.

Song, P. X.-K., Li, M., & Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics, 65*(1), 60–68.

Tekelab, T., Chojenta, C., Smith, R., & Loxton, D. (2019). Factors affecting utilization of antenatal care in Ethiopia: A systematic review and meta-analysis. *Plos One, 14*(4), e0214848.

Umberto Cherubin, Elisa Luciano, W. V. (2004). *Copula methods in finance*. New York: Wisely.

Warren, J. L., Perez-Heydrich, C., Burgert, C. R., & Emch, M. E. (2016). Influence of demographic and health survey point displacements on distance-based analyses. *Spatial Demography, 4*(2), 155–173.

WHO (2010). *Health Systems Financing: the path to universal Coverage (Arabic)*. New York: World Health Organization.

WHO (2016). *World health statistics 2016: monitoring health for the SDGs sustainable development goals*. New York: World Health Organization.

Yakoob, M. Y., Lawn, J. E., Darmstadt, G. L., & Bhutta, Z. A. (2010). Stillbirths: epidemiology, evidence, and priorities for action. In *Seminars in perinatology* (Vol. 34, pp. 387–394). Amsterdam: Elsevier.

Ye, Y., Yoshida, Y., Harun-Or-Rashid, M., & Sakamoto, J. (2010). Factors affecting the utilization of antenatal care services among women in Kham district, Xiengkhouang province, Lao PDR. *Nagoya Journal of Medical Science, 72*(1–2), 23–33.

Zelalem Ayele, D., Belayihun, B., Teji, K., and Admassu Ayana, D. (2014). Factors affecting utilization of maternal health care services in Kombolcha district, eastern Hararghe zone, Oromia regional state, Eastern Ethiopia. *International Scholarly Research Notices,* 917058, p. 7. https://doi.org/10.1155/2014/917058

Zhao, Q., Huang, Z. J., Yang, S., Pan, J., Smith, B., and Xu, B. (2012). The utilization of antenatal care among rural-to-urban migrant women in Shanghai: a hospital-based cross-sectional study. *BMC Public Health, 12*(1), 1012.

Part IV
Bayesian Statistical Modelling

Bayesian Survival Analysis with the Extended Generalized Gamma Model: Application to Demographic and Health Survey Data

Yuli Liang and Gebrenegus Ghilagaber

Abstract We extend the existing family of flexible survival models by assembling models scattered across the literature into a more knit form and under the same umbrella. New special cases are obtained not only by constraining the shape and scale parameters of the extended generalized gamma (EGG) model to fixed constants, but also by imposing relationships (such as equality, reciprocal, and negative reciprocal) between them. Apart from common parametric distributions such as exponential, Weibull, gamma, and log normal, the further extended family includes Rayleigh, inverse Rayleigh, ammag, inverse ammag, and half-normal distributions. The models are applied, in a Bayesian framework, on time to entry into first marriage among Eritrean men and women based on data from the 2010 Population and Health Survey. The application demonstrates that the further extended family of distributions provides a wide range of alternatives for a baseline distribution in the analysis of survival data. The empirical results reveal that the inverse gamma model fits best the data for men. It also performs closely as good as the EGG model in the data for women as well as in the combined sample.

Keywords Extended generalized gamma (EGG) model · Parametric survival models · Accelerated failure-time (AFT) models · Proportional hazards models · Bayesian inference · Parametric inference · Likelihood ratio test · Eritrea · Entry into marriage · Nuptiality · Demographic and Health Surveys (DHS) · Education · Residence · Exponential distribution · Inverse exponential distribution · Weibull distribution · Reciprocal Weibull distribution · Rayleigh distribution · Inverse Rayleigh distribution · Standard and generalized gamma

Y. Liang
School of Business and Economics, Department of Economics and Statistics, Linnaeus University, Växjö, Sweden
e-mail: Yuli.Liang@lnu.se

G. Ghilagaber (✉)
Department of Statistics, Stockholm University, Stockholm, Sweden
e-mail: Gebre@stat.su.se

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_11

distributions · Inverse gamma distribution · Ammag distribution · Inverse Ammag distribution · Log normal distribution · Half normal distributions · Log predictive density score (LPDS) · Censoring · Bayes Factors · Time to event data · Model comparison · Markov chain Monte Carlo (MCMC) · Random walk · Metropolis Hasting algorithm · Block sampling · Posterior distribution · Ergodic mean theorem · Inefficiency factor (IF)

1 Introduction

The usual goals in the analysis of survival data include: (a) describing the distributional shape of the time variable; (b) comparing the survival experiences of different groups in a population; and (c) modeling the relationship between explanatory variables and survival time—as measured by time to the event of interest or the rate at which the event occurs.

Two classes of models are common in the literature for investigating effects of explanatory variables on survival. In the Cox proportional hazards models, the explanatory variables act multiplicatively on a baseline hazard so that their effect is to increase or decrease the hazard relative to that of the baseline group. A second class of models, known as the accelerated failure-time models, specifies the covariates to act multiplicatively on time to event itself so that their effect is to accelerate or decelerate time to event relative to an event time for baseline group. According to Wei (1992), the accelerated failure-time model has an intuitive physical interpretation and would be a useful alternative to the Cox PH model in survival analysis.

It has been documented that covariate effects on survival time are not robust to the choice of the baseline distribution—see, for instance (Addison and Portugal, 1987; Bergström and Edin, 1992; Bergström et al., 1994; Ghilagaber, 2005). It is, therefore, of paramount importance to correctly specify the baseline distribution if results from analysis of survival data are to be utilized optimally. A number of distributions for survival data are available in the literature scattered across disciplines and application areas. Some previous works have attempted to put these scattered models in a more knit form by embedding a number of competing models under the umbrella of a general parametric framework as in Butler and McDonald (1986) and Peng et al. (1998). This enables the use of ordinary parametric inference for assessment of each competing model relative to a more comprehensive one. Among others, (Ghilagaber, 2005) shows that five parametric duration models (exponential, Weibull, gamma, log normal, and reciprocal Weibull) may be treated as special cases of a more general extended generalized gamma (EGG) model by constraining the shape and/or scale parameters of the EGG model to some fixed constants.

In this chapter, we extend the EGG model further and increase the family of flexible distributions to include 13 special cases. This is achieved by including distributions that not only constrain the shape and scale parameters to specified

constants but also impose some relationships between them. The new set of special cases include the Rayleigh and inverse Rayleigh distributions as well as the ammag and inverse ammag distributions as described in Cox et al. (2007). Further, a half-normal distribution can be obtained as a special case of ammag distribution.

A Bayesian approach is used to fit the EGG model and its 13 special cases to data on time to entry into first marriage among Eritrean men and women. Each special case model is then tested relative to a more general model using the log predictive density score (LPDS) in a Bayesian approach, see Li et al. (2010). Compared to the classical likelihood inference approach, the Bayesian approach provides three main advantages. First, we sample from a posterior density using Markov Chain Monte Carlo (MCMC), and hence, we can make exact inference for any sample size in any parametric survival models of various complexities. Second, we do not need to worry about the problem of local maximum trapping since our algorithm can go through the whole parameter spaces supported by the data. Third, it is straightforward to investigate the performance of joint posterior density, whereas in a frequentist paradigm, we need to run simulation by pre-specifying the true values of parameters when evaluating the performance of maximum likelihood estimates.

In Sect. 2, we introduce the accelerated failure-time models and demonstrate how a number of common distributions can be brought under the umbrella of the EGG model. Bayesian density estimation of the EGG model and MCMC implementation is described in Sect. 3. In Sect. 4, we illustrate the models of Sect. 2 and the methods of Sect. 3 using real-life data from the 2010 Eritrean Population and Health Survey. Section 5 concludes the chapter by way of summary and concluding remarks. A full list of the distributions used in this chapter, a proof for a lemma, and the R code used in the illustrative example are provided in Appendices.

2 Parametric Models for Survival Data

2.1 Background

Survival data contain information on durations until event or censoring $(t_1, t_2, ..., t_n)$ together with a censoring indicator as well as background variables or covariates $(z_1, z_2, ..., z_p)$ that are often socio-demographic characteristics of individuals or organizations. The distribution of survival time, T, may be described by its three equivalent functions: the survival function, $S(t) = P(T > t)$, the density function, $f(t)$, or the hazard (intensity) function, $h(t) = f(t)/S(t)$, where the last two functions require absolute continuity.

These functions can vary not only over time, but also among individuals within a population. Thus, one objective in the analysis of survival data is to draw inferences about the influence of covariates on these functions. One popular model is the Cox proportional hazards model presented in Cox (1972) where a p-dimensional vector of covariates $\mathbf{z}$ affects the hazard function in a multiplicative manner according to

$$h(t|\mathbf{z}) = h_0(t) \exp\left(\mathbf{z}'\boldsymbol{\beta}\right), \tag{1}$$

where $h_0(t)$ is an unspecified baseline function of time and $\boldsymbol{\beta} \in R^p$ is an unknown vector of parameters representing the effect of the covariates $\mathbf{z}$. The factor $\exp(\mathbf{z}'\boldsymbol{\beta})$ describes the intensity (hazard) for an individual with vector $\mathbf{z}$ relative to that of a standard individual (with $\mathbf{z} = \mathbf{0}$).

2.2 Accelerated Failure-Time Models

A second class of models, the accelerated failure-time model, specifies the covariates to act multiplicatively on the event time itself rather than on the hazard function.

If T_0 is the random time to event associated with an individual in the baseline group ($\mathbf{z} = \mathbf{0}$), then the accelerated failure-time model specifies that for an individual with a non-zero vector of covariates $\mathbf{z}$, the event time is given by

$$T = T_0 \exp(\mathbf{z}'\boldsymbol{\beta}) \tag{2}$$

or equivalently

$$\ln(T) = \mathbf{z}'\boldsymbol{\beta} + \ln(T_0), \tag{3}$$

where, as before, T is the event time, $\mathbf{z}$ is a vector of covariates, and $\boldsymbol{\beta}$ is a vector of regression parameters. Since covariates alter, by a scale factor, the rate at which an individual traverses the time axis, Eq. (2) is referred to as the accelerated failure-time model. Thus, in accelerated failure-time models, the effect of the explanatory variables is to accelerate or decelerate time to event relative to T_0.

The model in (3) is a linear model with $\ln(T_0)$ playing the role of an error term with an underlying baseline distribution. Usually, a scale parameter δ is allowed in the model to give

$$\ln(T) = \mathbf{z}'\boldsymbol{\beta} + \delta \ln(T_0) = \mathbf{z}'\boldsymbol{\beta} + \delta\epsilon, \tag{4}$$

where a more conventional notation ϵ is used for the error term.

From (4), we note that $T = e^{\mathbf{z}'\boldsymbol{\beta}} T_0^{\delta}$. Thus, the survival function of T may be written in terms of that of T_0:

$$S(t) = P(T > t) = P(e^{\mathbf{z}'\boldsymbol{\beta}} T_0^{\delta} > t) = P(T_0^{\delta} > te^{-\mathbf{z}'\boldsymbol{\beta}}) = S_0(te^{-\mathbf{z}'\boldsymbol{\beta}}), \tag{5}$$

where $S_0(.)$ is the survival function of the baseline time with scale parameter δ, T_0^{δ}, and $e^{-\mathbf{z}'\boldsymbol{\beta}}$ is the accelerating/decelerating factor. In other words, the probability for an individual with covariate vector $\mathbf{z}$ surviving beyond time t is the same as the probability for an individual in the baseline group ($\mathbf{z} = \mathbf{0}$) surviving beyond time

$te^{-\mathbf{z}'\boldsymbol{\beta}}$. A positive coefficient $\boldsymbol{\beta}$ shifts the time $te^{-\mathbf{z}'\boldsymbol{\beta}}$ to the left of t, while a negative $\boldsymbol{\beta}$ shifts the time $te^{-\mathbf{z}'\boldsymbol{\beta}}$ to the right of t if all components of $\mathbf{z} > 0$. Accordingly, the density and hazard functions can also be written in terms of the baseline density and hazard:

$$f(t) = e^{-\mathbf{z}'\boldsymbol{\beta}} f_0(te^{-\mathbf{z}'\boldsymbol{\beta}})$$
$$h(t) = e^{-\mathbf{z}'\boldsymbol{\beta}} h_0(te^{-\mathbf{z}'\boldsymbol{\beta}}).$$

The distribution of T_0 in (4) may be selected from positive-valued distributions such as Weibull or log normal that, in turn, yield extreme-value and normal distributions for the error term ϵ. Below, we demonstrate how the list may be expanded by assembling various models under the same umbrella.

2.3 The Extended Generalized Gamma (EGG) Model

Stacy (1962) introduced the generalized gamma model that is useful in embedding competing models into a single parametric framework. This model is the distribution of T such that $\ln(T) = \mu + \delta\epsilon$, where $\mu \in R$, $\delta > 0$, and the random error term ϵ has the density

$$f(k, \epsilon) = \frac{1}{\Gamma(k)} \exp\left[k\epsilon - \exp(\epsilon)\right], k > 0,$$

where k is an additional shape parameter. Prentice (1974) showed that a shift of parameter of the form $q = k^{-\frac{1}{2}}$ leads to a standard normal distribution for T giving an interior point for $q = 0$ in the parameter space. The final model with parameters $\mu, q \in R$ and $\delta > 0$ can be written as $\ln(T) = \mu + \delta\epsilon$, where the error density function $f(q, \epsilon)$ is given by

$$f(q, \epsilon) = \begin{cases} \frac{|q|}{\Gamma(q^{-2})}(q^{-2})^{q^{-2}} \exp\left\{q^{-2}\left[q\epsilon - \exp(q\epsilon)\right]\right\}, & q \neq 0 \\ \frac{1}{\sqrt{2\pi}} \exp(-\frac{\epsilon^2}{2}), & q = 0. \end{cases} \tag{6}$$

The distribution of T when the error term has the density given in Eq. (6) is known as the extended generalized gamma (EGG) distribution, see, for instance (Ghilagaber, 2005; Ghilagaber et al., 2014).

As can be seen from the lower part of (6), the EGG model reduces to the standard normal distribution for ϵ when the shape parameter q is equal to zero. Accordingly, T will have a log-normal distribution. When the shape parameter $q = 1$, (6) reduces to

$$f(q,\epsilon) = \exp\left[\epsilon - \exp(\epsilon)\right],$$

which is the standard (type 1) extreme-value distribution. As $\ln(T)$ is a linear function of ϵ, it has the same (extreme-value) distribution as ϵ. Hence, $T = \exp(\mathbf{z}'\boldsymbol{\beta} + \delta\epsilon)$ as defined in Eq. (4) will have a Weibull distribution. If $q = 1$ and $\delta = 1$, then T has the exponential distribution as a special case of the Weibull distribution. The case of $q = -1$ corresponds to extreme maximum-value distribution for $\ln(T)$. This, in turn, corresponds to reciprocal Weibull distribution for T.

The case of $\delta = 1$ and $q > 0$ is also of interest. Farewell and Prentice (1977) argue that this gives the ordinary gamma distribution for T. Others, (Bergström and Edin, 1992; Bergström et al., 1994, 1997), argue that this did not hold in their case illustrations. Consequently, we shall relax this special case to $\delta = 1$ and $q \in R$ and label it the "gamma" distribution in our illustrative example. Below, we further extend the above family of distributions by imposing some relationships between the scale and shape parameters.

2.4 Further Extensions of the EGG Model

We begin with a baseline distribution for time to event, $T_0 \sim EGG(0,1,q)$, and label it as standard generalized gamma distribution with density and survival functions given by

$$f_{EGG(0,1,q)}(t_0) = \frac{|q|}{t_0\Gamma(q^{-2})}(q^{-2}t_0^q)^{q^{-2}}exp(-q^{-2}t_0^q),$$

$$S_{EGG(0,1,q)}(t_0) = \begin{cases} 1-\Phi(\ln t_0), & q=0 \\ 1-\gamma(q^{-2}, t_0^q q^{-2})/\Gamma(q^{-2}), & q>0 \\ \gamma(q^{-2}, t_0^q q^{-2})/\Gamma(q^{-2}), & q<0, \end{cases}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. By transformation, $t = e^{\mu}t_0^{\delta} \sim EGG(\mu,\delta,q)$, and T is said to have the extended generalized gamma distribution with shape parameter $\mu \in R$, scale parameter $\delta > 0$, and an additional index shape parameter $q \in R$. We denote this by $T \sim EGG(\mu,\delta,q)$, with density

$$f_{EGG}(t) = \frac{|q|}{t\delta\Gamma(q^{-2})}\left[q^{-2}(e^{-\mu}t)^{\frac{q}{\delta}}\right]^{q^{-2}} exp\left[-q^{-2}(e^{-\mu}t)^{\frac{q}{\delta}}\right] \qquad (7)$$

$$= \begin{cases} \frac{1}{\delta t\sqrt{2\pi}} e^{-\frac{(\ln t-\mu)^2}{2\delta^2}} & q = 0 \\ \frac{q}{\delta} t^{\frac{q}{\delta}-1} \frac{1}{\Gamma(q^{-2})} \left[q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]^{q^{-2}} (t^{\frac{q}{\delta}})^{q^{-2}-1} exp\left[-q^{-2}(e^{-\mu})^{\frac{q}{\delta}} t^{\frac{q}{\delta}}\right] & q > 0 \\ -\frac{q}{\delta} t^{\frac{q}{\delta}-1} \frac{1}{\Gamma(q^{-2})} \left[q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]^{q^{-2}} (t^{\frac{q}{\delta}})^{q^{-2}-1} exp\left[-q^{-2}(e^{-\mu})^{\frac{q}{\delta}} t^{\frac{q}{\delta}}\right] & q < 0. \end{cases}$$

The component

$$\frac{1}{\Gamma(q^{-2})} \left[q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]^{q^{-2}} (t^{\frac{q}{\delta}})^{q^{-2}-1} exp\left[-q^{-2}(e^{-\mu})^{\frac{q}{\delta}} t^{\frac{q}{\delta}}\right]$$

in the above equation is the density of the gamma distribution for $t^{\frac{q}{\delta}}$ with a shape parameter q^{-2} and a rate parameter $q^{-2}(e^{-\mu})^{\frac{q}{\delta}}$. The next lemma gives the rth moment and the first four central moments of the EGG density. The following definitions of skewness and excess kurtosis are used:

$$S(T) = \frac{E\left[T - E(T)\right]^3}{V(T)^{3/2}},$$

$$K(T) = \frac{E\left[T - E(T)\right]^4}{V(T)^2} - 3,$$

where $V(T)$ is the variance.

Lemma 1 *If* $T \sim EGG(\mu, \delta, q)$, *then*

$$E(T^r) = \begin{cases} \frac{\Gamma\left(q^{-2}+r\frac{\delta}{q}\right)}{\left(q^{\frac{-2\delta}{q}} e^{-\mu}\right)^r \Gamma(q^{-2})}, & \text{if } r\delta/q > -q^{-2}, \\ \infty & \text{otherwise.} \end{cases}$$

$$E(T) = \frac{\Gamma(q^{-2} + \frac{\delta}{q})}{q^{\frac{-2\delta}{q}} e^{-\mu} \Gamma(q^{-2})},$$

$$V(T) = \frac{\Gamma(q^{-2} + \frac{2\delta}{q})\Gamma(q^{-2}) - \Gamma^2(q^{-2} + \frac{\delta}{q})}{\Gamma^2(q^{-2})(q^{\frac{-2\delta}{q}} e^{-\mu})^2},$$

$$E\left[T - E(T)\right]^3 = 2\Gamma^3\left(q^{-2} + \frac{\delta}{q}\right) - 3\Gamma\left(q^{-2} + \frac{2\delta}{q}\right)\Gamma\left(q^{-2} + \frac{\delta}{q}\right)\Gamma(q^{-2})$$

$$+\Gamma\left(q^{-2} + \frac{3\delta}{q}\right)\Gamma^2(q^{-2}),$$

$$E\left[T-E(T)\right]^4 = -3\Gamma^4\left(q^{-2}+\frac{\delta}{q}\right)+6\Gamma\left(q^{-2}+\frac{2\delta}{q}\right)\Gamma^2\left(q^{-2}+\frac{\delta}{q}\right)\Gamma(q^{-2})$$
$$-4\Gamma\left(q^{-2}+\frac{3\delta}{q}\right)\Gamma\left(q^{-2}+\frac{\delta}{q}\right)\Gamma^2(q^{-2})+\Gamma\left(q^{-2}+\frac{4\delta}{q}\right)\Gamma^3(q^{-2}).$$

A simplified proof of Lemma 1 is provided in Appendix 2.

From Lemma 1, we note that $S(T)$ and $K(T)$ are the functions of q and δ/q, implying both q and δ/q are shape parameters.

The survival function of t is then given by

$$S_{EGG}(t) = \begin{cases} 1-\Phi\left(\frac{\ln t-\mu}{\delta}\right) & q=0 \\ 1-\gamma\left[q^{-2}, t^{\frac{q}{\delta}}q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]/\Gamma(q^{-2}) & q>0 \\ \gamma\left[q^{-2}, t^{\frac{q}{\delta}}q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]/\Gamma(q^{-2}) & q<0, \end{cases} \tag{8}$$

where $\gamma\left[q^{-2}, t^{\frac{q}{\delta}}q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]/\Gamma(q^{-2})$ is the corresponding cumulative distribution function of the gamma distribution for $t^{\frac{q}{\delta}}$ when $q>0$ and $\gamma\left[q^{-2}, t^{\frac{q}{\delta}}q^{-2}(e^{-\mu})^{\frac{q}{\delta}}\right]$ is a lower incomplete gamma function with the form of $\gamma(s,r)=\int_0^r x^{s-1}e^{-x}dx$ described in Abramowitz and Stegun (1964).

The EGG model redefined in Eqs. (7) and (8) is a rich and versatile model containing many special cases based on different combinations of q and δ.

Apart from those mentioned in the previous subsection, the list may be extended to include the inverse exponential ($q=-\delta=-1$), standard gamma ($q=\delta$), inverse gamma (when $q=-\delta$), ammag ($q=1/\delta$), inverse ammag ($q=-1/\delta$), Rayleigh ($q=1$ and $\delta=1/2$), inverse Rayleigh ($q=-1$ and $\delta=1/2$), and half-normal ($q=\sqrt{2}$ and $\delta=\sqrt{2}/2$).

EGG nests more special cases such as Maxwell–Boltzmann, but we have not included this in the present chapter since our focus is on the distribution of survival time T. Further, the equivalent distributions of some special cases are excluded. For instance, the inverse gamma model is equivalent to the Levy model in some special cases: inverse gamma$(q^{-2}, q^{-2}e^{-\mu}) \leftrightarrow Levy(0,c)$ when $q^{-2}=1/2$ and $q^{-2}e^{-\mu}=2c$. The standard gamma model is also equivalent to a chi-squared model in some situations: standard gamma$(q^{-2}, q^{-2}e^{-\mu}) \leftrightarrow \chi^2_{(v)}$ when $q^{-2}=v/2$ and $q^{-2}e^{-\mu}=1/2$.

To sum up, the EGG model constitutes of at least 13 special cases whose relationships are depicted in Fig. 1. Each special case model can be assessed relative to a more comprehensive one using appropriate procedures for comparing nested models. A summary of the density functions, $f(t)$, and survival functions, $S(t)$, for 13 special cases is provided in Appendix 1. The corresponding hazard functions can be obtained by $h_{EGG}(t)=f_{EGG}(t)/S_{EGG}(t)$. The hazards in the EGG models

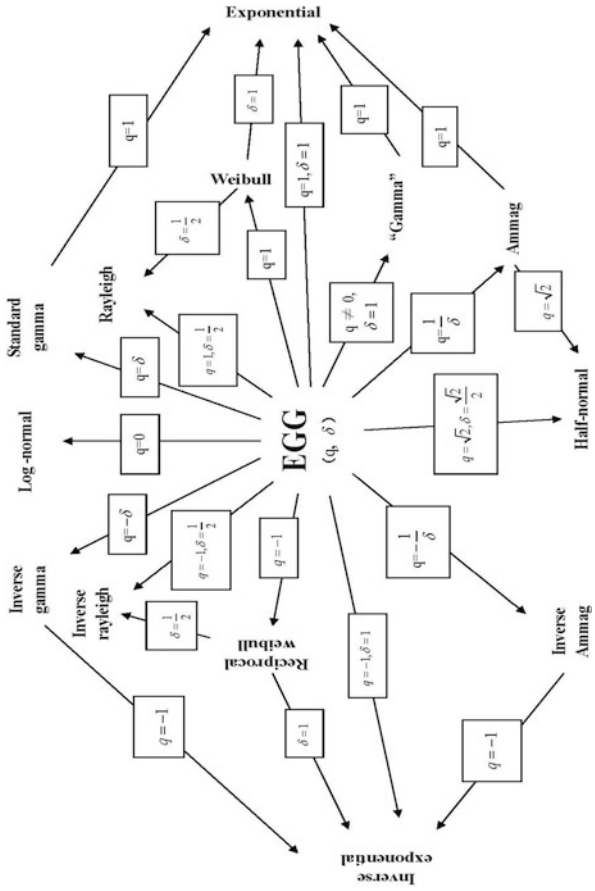


Fig. 1 Relationships among EGG family of distributions

can be of various forms—increasing, decreasing, bathtub, or arc-shaped (Cox et al., 2007).

When we adapt the generalized gamma distribution to accelerated failure-time models, the location parameter μ can be composed of a linear predictor based on p covariates $\mu = \beta_0 + \sum_{i=1}^{n} z_{ji}\beta_j$ $(j = 1 \cdots p)$, which justifies the feasibility of the EGG in accelerated failure-time models.

The distribution of $\epsilon = \ln(T_0)$ is given in Eq. (6). When $q = 0$, ϵ is standard normal distributed; when $q \neq 0$, it can be manipulated to give

$$f(\epsilon; q) = |q|\, e^{q\epsilon} \frac{(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} (e^{q\epsilon})^{q^{-2}-1} exp\left[-q^{-2} exp(q\epsilon)\right] \tag{9}$$

with the corresponding survival functions

$$S(\epsilon, q) = \begin{cases} 1 - \Phi(\epsilon) & q = 0 \\ 1 - \gamma\left[q^{-2}, \exp(q\epsilon)q^{-2}\right] / \Gamma(q^{-2}) & q > 0 \\ \gamma\left[q^{-2}, \exp(q\epsilon)q^{-2}\right] / \Gamma(q^{-2}) & q < 0. \end{cases} \tag{10}$$

Based on the density of ϵ, Fig. 2 shows the shape of some density functions, $f(\epsilon)$, for some selected values of q. Here, we have a special case of $\ln(T) = \mathbf{z}'\boldsymbol{\beta} + \delta\epsilon = \mu + \delta\epsilon$

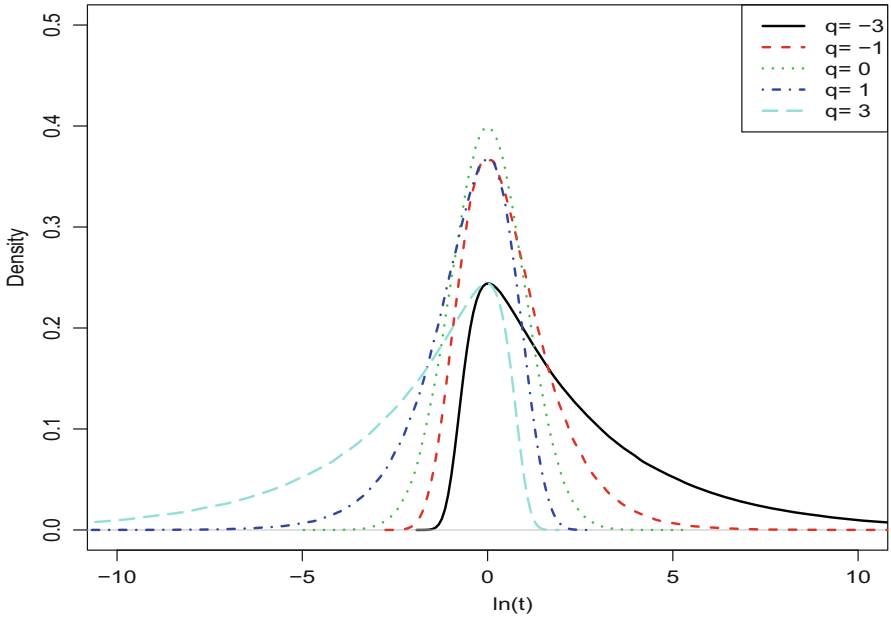


Fig. 2 Five distributions of $\ln(T)$ for $\mu = 0$, $\delta = 1$, and some values of q

in which $\mu = 0$ and $\delta = 1$. We note that the densities are positively skewed for $q < 0$ and negatively skewed for $q > 0$ with both the absolute skewness and kurtosis monotone increasing in $|q|$—which are in accordance with those of Prentice (1974).

3 Bayesian Inference in the Extended Generalized Gamma Model

Bayesian inference for a three-parameter EGG model and four-parameter generalized gamma distribution (EGG model with one extra location parameter) is discussed in Tsionas (2001) and Van Noortwijk (2001) for situations where there is no censoring. Inference becomes more complicated in the presence of censored observations due to, for instance, difficulty to find conjugate prior or derive full conditional posterior.

Heleno and Alberto (1986) have used Bayesian approach for EGG model with censored data using Jeffrey multi-parameter prior. Ramos et al. (2017) have shown that both the Jeffreys prior and the reference priors give improper posteriors to the EGG model, and then proposed the overall reference prior in Berger et al. (2015), which provided the proper posterior. In this section, we present Bayesian inference in the EGG model that allows for any type of censoring mechanism.

3.1 Prior and Posterior

In a Bayesian framework, any prior information about the parameters of interest is combined with the data (likelihood) to derive a posterior distribution.

In our present case, we use normal priors with mean 0 and large variance σ_1^2 for each effect parameters $\beta_j (j = 0, \cdots, p)$. We also assume a vague prior, a gamma distribution with hyperparameters a and b for the scale parameter δ. For the index shape parameter q, a normal prior with mean 0 and large variance σ_2^2 is assumed. These independent priors can be summarized as follows:

$$\begin{aligned}
&\beta_j \sim N(0, \sigma_1^2), \;\; j = 1, \ldots, p \\
&\delta \sim Gamma(a, b) \\
&q \sim N(0, \sigma_2^2).
\end{aligned}$$

We can use any prior that reflects our prior knowledge (if any) of the unknown parameters. In our illustration in Sect. 4, we will use $\sigma_1 = \sigma_2 = 1000$ and hyperparameters $a = b = 1$. The rationale behind this is to let the likelihood dominate the posterior so that the inferences drawn are driven by the data.

Denoting data with $\mathcal{D}$, the joint posterior distribution is then given by

$$f(\boldsymbol{\beta}, \delta, q|\mathcal{D}) \propto L(\boldsymbol{\beta}, \delta, q; \mathcal{D}) f(\boldsymbol{\beta}, \delta, q)$$
$$= L(\boldsymbol{\beta}, \delta, q; \mathcal{D}) \prod_{j=1}^{p} f(\beta_j) f(\delta) f(q),$$

where $L(\boldsymbol{\beta}, \delta, q; \mathcal{D})$ is the likelihood function, and $f(\cdot)$ is the prior density function of β_j, δ, and q with known hyperparameters. The above posterior can be generalized to other types of likelihood functions based on other censoring mechanisms (than the standard right censoring assumed in our present case). With right censored data, the likelihood function becomes

$$L(\boldsymbol{\beta}, \delta, q; \mathcal{D}) = \prod_{i=1}^{n} (\delta^{-1} f(\epsilon_i, q))^{d_i} S(\epsilon_i, q)^{1-d_i},$$

where d_i is the censoring indicator and $f(\epsilon_i, q)$ and $S(\epsilon_i, q)$ are given by Eqs. (9) and (10), respectively.

Since there is no explicit analytical form for the posterior distribution, sampling is performed using numerical methods based on Markov Chain Monte Carlo (MCMC).

3.2 *MCMC: Random Walk Metropolis–Hastings Algorithm with Block Sampling*

We sample all parameters sequentially from the joint posterior distribution using the Metropolis–Hastings algorithm. See Gelman et al. (2004) for more details on the Metropolis–Hastings algorithm and its nice properties. A random walk Metropolis–Hastings algorithm with block sampling is used, and the sampling procedure for the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \delta, q)'$ can be summarized as follows:

(1) Set the initial values for the parameters $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \delta_0, q_0)'$.
(2) Construct the proposal distribution $J(\boldsymbol{\theta}_p|\boldsymbol{\theta}_c) \sim N(\boldsymbol{\theta}_c, c^2\boldsymbol{\Sigma})$, where $\boldsymbol{\theta}_p$ is the candidate value, $\boldsymbol{\theta}_c$ is the current value, and c is the scaling constant and $\boldsymbol{\Sigma}$ is a known covariance matrix. Here we choose $\boldsymbol{\Sigma} = -H^{-1}(\widehat{\boldsymbol{\theta}})$, where $H(\widehat{\boldsymbol{\theta}})$ is the Hessian matrix evaluated at $\widehat{\boldsymbol{\theta}}$, which is obtained by Newton's method. Following Gelman et al. (2004), we choose a value of $c = 2.4/\sqrt{k}$, where k is the length of the vector $\boldsymbol{\theta}$.
(3) Generate $\boldsymbol{\theta}^*$ from $J(\boldsymbol{\theta}_p|\boldsymbol{\theta}_c)$ and U from $U(0, 1)$.
(4) If

$$U < \frac{f(\boldsymbol{\theta}^*|\mathcal{D}) f(\boldsymbol{\theta}^*) J(\boldsymbol{\theta}_c|\boldsymbol{\theta}_p)}{f(\boldsymbol{\theta}_c|\mathcal{D}) f(\boldsymbol{\theta}_c) J(\boldsymbol{\theta}_p|\boldsymbol{\theta}_c)},$$

the candidate vector $\boldsymbol{\theta}^*$ is accepted and $\boldsymbol{\theta}_c = \boldsymbol{\theta}^*$; otherwise, we keep $\boldsymbol{\theta}_c$.
(5) Return to step (2).

3.3 Posterior Statistics and Convergence Diagnostics

We summarize our posterior distribution by way of posterior means and highest posterior density (hpd). Since MCMC is based on ergodic mean theorem (Markov chain law of large numbers), convergence can be verified using diagnostic plots such as a plot of the cumulative mean against the number of iterations. In addition, inefficiency factors (IF) can be computed as a measure of the efficiency of the MCMC scheme.

3.4 Bayesian Model Comparisons

The common way of comparing models in the Bayesian framework is the use of Bayes factor that is the ratio of marginal likelihood of two competing models.

Suppose we have a set of candidate models $\mathcal{M}_m, m = 1, \cdots, M$ and the corresponding model parameters $\boldsymbol{\theta_m}$. The posterior model probability is then given by

$$P(\mathcal{M}_m|Y) \propto P(\mathcal{M}_m)P(Y|\mathcal{M}_m),$$

where Y represents the data at hand. The posterior odds $P(\mathcal{M}_m|Y)/P(\mathcal{M}_l|Y)$ can be used to compare two models, and it can be written in terms of the Bayes factor:

$$\frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} BF_{ml},$$

where BF_{ml} is the Bayes factor between $\mathcal{M}_m$ and $\mathcal{M}_l$ with the form

$$BF(Y) = \frac{P(Y|\mathcal{M}_m)}{P(Y|\mathcal{M}_l)} = \frac{\int P(Y|\boldsymbol{\theta_m}, \mathcal{M}_m)P(\boldsymbol{\theta_m}|\mathcal{M}_m)d\boldsymbol{\theta_m}}{\int P(Y|\boldsymbol{\theta_l}, \mathcal{M}_l)P(\boldsymbol{\theta_l}|\mathcal{M}_l)d\boldsymbol{\theta_l}}.$$

The marginal likelihood is a conditional expectation for the likelihood given the prior

$$E_{P(\boldsymbol{\theta_m}|\mathcal{M}_m)}(P(Y|\boldsymbol{\theta_m}, \mathcal{M}_m)).$$

It is sensitive to the choice of the prior, especially when the prior is not very informative (Villani et al., 2009). For instance, if $P(\boldsymbol{\theta_m}|\mathcal{M}_m)$ is far from $P(Y|\boldsymbol{\theta_m}, \mathcal{M}_m)$, while $P(\boldsymbol{\theta_l}|\mathcal{M}_l)$ is close to $P(Y|\boldsymbol{\theta_l}, \mathcal{M}_l)$, it is possible that $P(Y|\mathcal{M}_m)$ is less than $P(Y|\mathcal{M}_l)$ even though $\mathcal{M}_l$ is a sub-model of $\mathcal{M}_m$.

To avoid such sensitivity to the choice of prior, we compare our models in the illustration on the basis of their predictive performance. The data is split randomly into B folds, and B-1 fold is used as a training data $\tilde{y}_{-b}$, while the rest one-fold is

used as a testing data $\tilde{y}_b$. The B-fold cross-validation of the log predictive density score (LPDS) is then formed as

$$B^{-1}\sum_{b=1}^{B}\ln p(\tilde{y}_b|\tilde{y}_{-b},x).$$

In other words, part of the observations are used to update the flat (non-informative) prior and the sensitivity to the prior can be reduced substantially. According to Villani et al. (2009), the Bayes factor is roughly B times more discriminatory than the LPDS. For selecting models in Sect. 4, the LPDS was computed using B = 5 folds of the data.

4 Application: Educational and Residential Differences in Marriage Timing Among Eritrean Men and Women

We now illustrate the models and methods described in the previous sections with real-life data—entry into marriage among Eritrean men and women based on its 2010 Population and Health Survey (EPHS2010).

The main goals with the illustration are to study the distributional shapes of the times to marriage, model the effects of covariates on these event times, and examine if inferences regarding covariate effects are robust to the choice of distributional shape.

The study of marriage timing (age at marriage) is also of substantive interest in its own because of its strong negative association with women's health directly (Raj, 2010) or indirectly through its negative impact on health care utilization (Godha et al., 2016).

4.1 Data and Variables

The data used for illustration in this chapter come from the 2010 Eritrea Population and Health Survey, EPHS2010 (National-Statistics-Office-Eritrea and Fafo-AIS, 2013). The EPHS2010 was designed as a follow-up to its predecessors—the 1995 and 2002 Demographic and Health Surveys (National-Statistics-Office-Eritrea and Macro-International-Inc., 1997, 2003), and to update the information from the previous surveys as well as provide findings on some new topics of interest.

The EPHS2010 was conducted between January and July 2010 and gathered information from 30224 women aged 15–49 and 5021 men aged 15–59. For the purpose of this paper, only respondents with known values on marital status at the time of the survey are used in the analyses. This resulted in 10238 usable records for women and all 5021 records for men. Detailed tabulations for the entire survey

Table 1 Summary statistics of the data sets used in the illustration

		Women			Men			Combined Sample		
Covariate	Levels	n	Events	% Event	n	Event	% Event	n	Event	% Event
	No Educ	4186	3799	90.75	1050	892	84.95	5236	4691	89.59
	Primary	2055	1634	79.51	803	543	67.62	2858	2177	76.17
Education	Middle	1827	1006	55.06	1209	455	37.63	3036	1461	48.12
	Secon	1894	886	46.78	1516	461	30.41	3410	1347	39.50
	PostSec	276	96	34.78	442	218	49.32	718	314	43.73
	Capital	1819	969	53.27	931	344	36.95	2750	1313	47.75
Residence	Other Towns	2504	1739	69.45	1257	565	44.95	3761	2304	61.26
	Rural Areas	5915	4713	79.68	2833	1660	58.60	8748	6373	72.85
	Total	10,238	7421	72.49	5021	2569	51.17	15,259	9990	65.47

may be found in the EPHS2010 Final Report (National-Statistics-Office-Eritrea and Fafo-AIS, 2013). Summary statistics for the subset of data used in the present chapter are shown in Table 1.

By the survey time (January–July 2010), 7421 of the 10238 women (72 %) and 2569 of the 5021 men (51 %) have responded they were ever married (this includes those who might have been separated or widowed after). The rest, 2817 women and 2452 men (28 % and 49 %, respectively), have responded that they were still single at the time of interview. The distribution of the women across educational level shows that 4186 (41 %) had no education at all, 2055 (20 %) had primary-level education, 1827 (18 %) had middle-level education, 1894 (18 %) had secondary-level education, while the rest 276 (3 %) had post-secondary education. The corresponding figures for men are 1051 (21 %), 803 (16 %), 1209 (24 %), 1516 (30 %), and 442 (9 %), respectively. Further, 1819 (18 %) of the women respondents were from the capital (Asmara), 2504 (24 %) were from other towns, while the majority 5915 (58 %) were from rural areas. The corresponding figures for men are 931 (19 %), 1257 (25 %), and 2833 (56 %), respectively.

The columns of percentage married in Table 1 reveal clear differentials across both educational levels and residence for both women and men. For instance, while women with no education constitute 41 % of the entire sample, they constitute 51 % of the marriages (3799 of 7421). Women with post-secondary education, on the other hand, constitute only 1 % of the marriages (96 of 7421). The pattern is similar but less dramatic for men—those with no education constitute 35 % of the marriages, while those with post-secondary education constitute only 8 % of the marriages. Differentials across residence show that women from rural areas constitute 58 % of the sub-sample but 64 % of the marriages. Women from the capital, on the other hand, constitute 19 % of the sub-sample but only 13 % of the marriages. The contribution of men from the capital to the sub-sample is 18 %, while their contribution to the total marriage is 15 %. Men from rural areas constitute 56 % of the sub-sample but 65 % of the marriages.

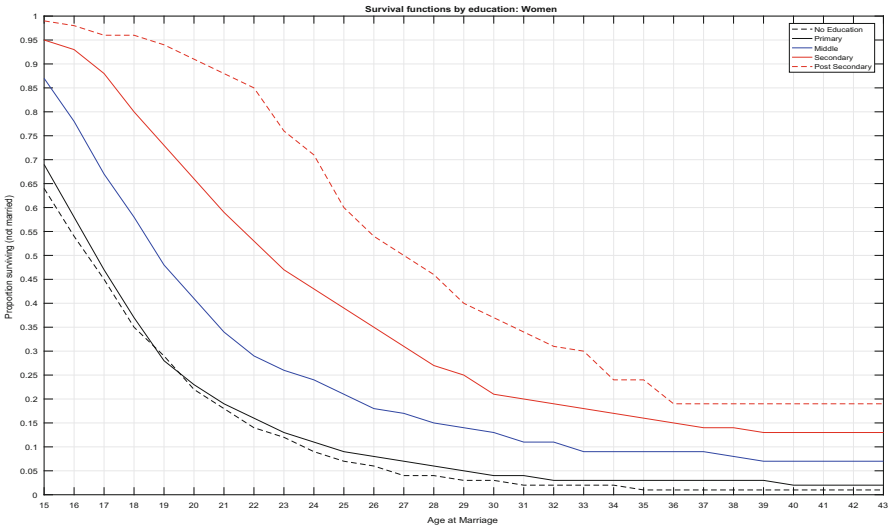


Fig. 3 Survival functions by education: Women

Plots of survival functions by education and residence for women and men are shown in Figs. 3, 4, 5, 6, and 7. Figures 3 and 4 show plots for women by education and residence, respectively, while Figs. 5 and 6 show the corresponding plots for men. Figure 7 shows gender differences in entry to first marriage among all men and women.

The plots depict what we already noted in Table 1—that there are differentials across education and residence and that the educational differences are more pronounced in the women data than in men data. The last figure shows that women enter marriage at faster rates than men.

The summary in Table 1 and Figs. 3, 4, 5, 6, and 7 provides a good description of the data at hand, but in order to make sound inferences based on the sub-sample, we need deeper analyses of the data and formal statistical tests. Ghilagaber (2018) has analyzed the data sets using frequentist statistical methods ranging from elementary measures of association between an event of interest and background variables to more complex and advanced methods that utilize the data more efficiently. Elsewhere in this book, (Munezero and Ghilagaber, 2022b) analyze the data sets using dynamic Bayesian approach where covariate effects are allowed to vary over time.

In the next sub-section, we present and discuss results from fitting the further EGG model of Sect. 2 to the above data sets in the Bayesian framework of Sect. 3.

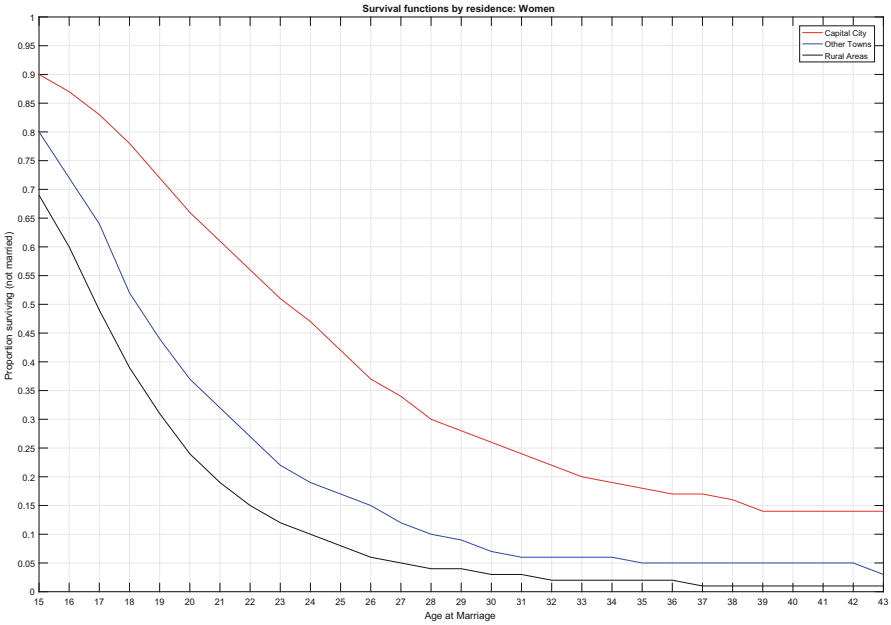


Fig. 4 Survival functions by residence: Women

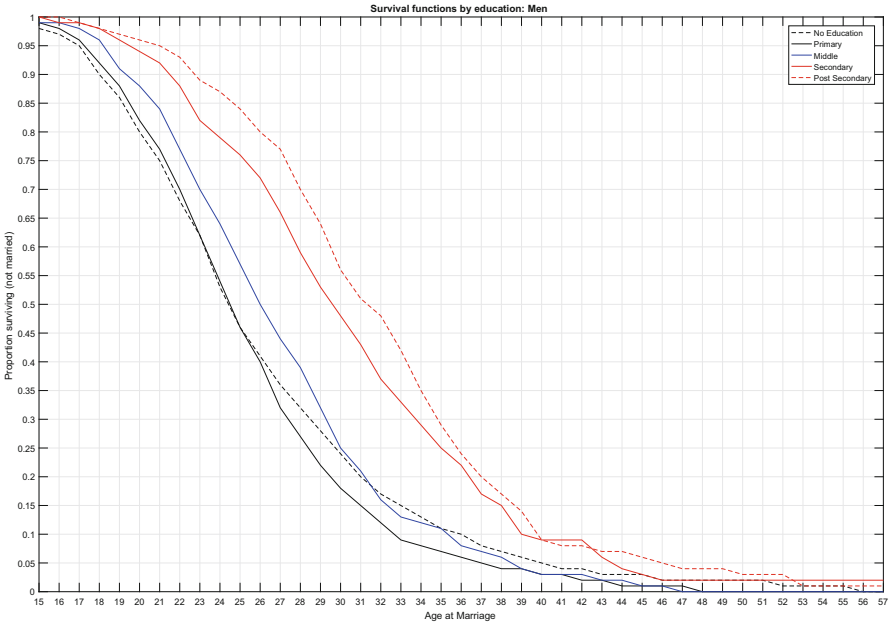


Fig. 5 Survival functions by education: Men

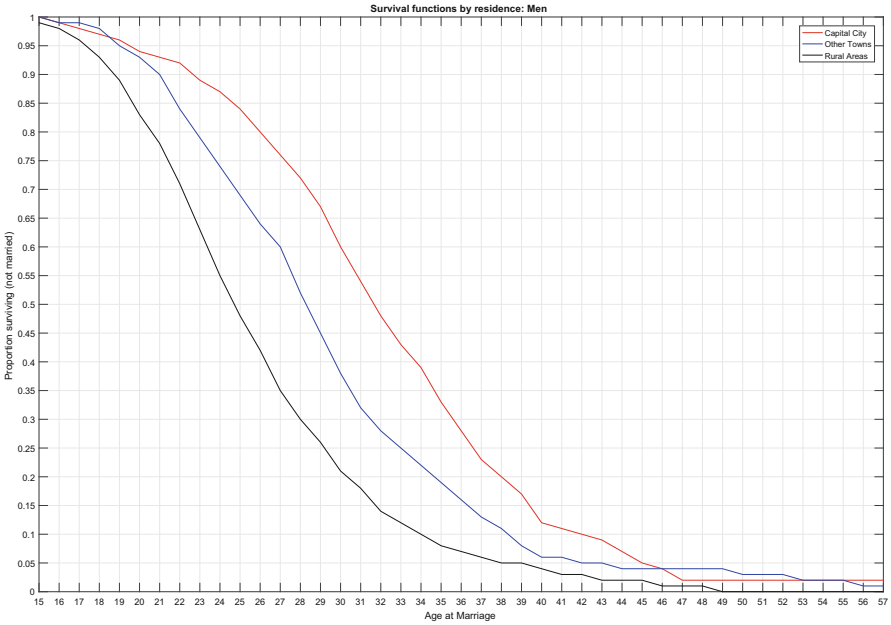


Fig. 6 Survival functions by residence: Men

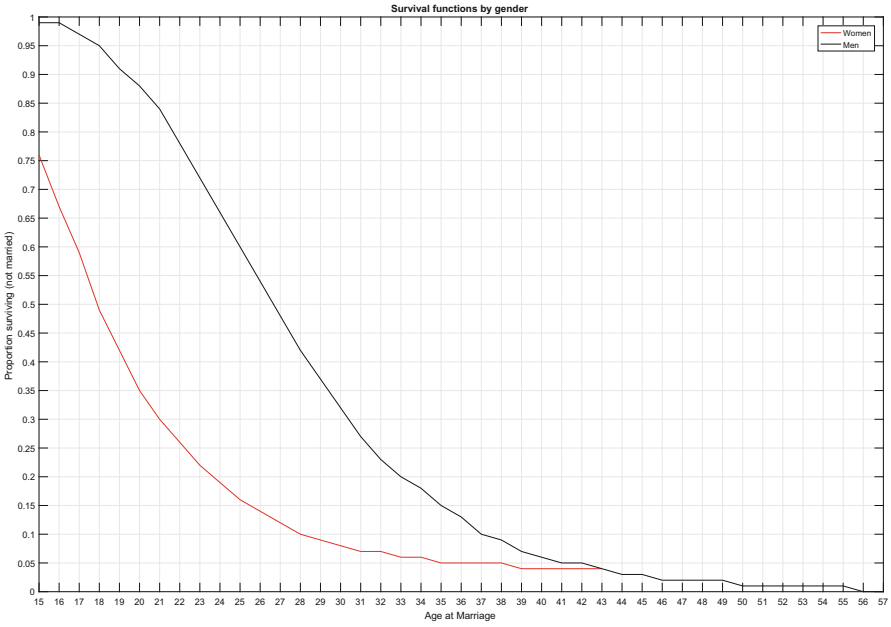


Fig. 7 Survival functions by gender

4.2 *Results from Bayesian Analysis of the Data Using the EGG Model*

Table 2 contains a summary of our results to which we will return at the end of this section. Results from fitting the extended generalized gamma (EGG) model and its 13 special cases to the data for women, men, and the combined sample are shown in Tables 3, 4, and 5, respectively.

In Table 3, the results from the unconstrained EGG model show that the scale and shape parameters (which are freely estimated from the data) are $\delta = 0.246$ and $q = -0.526$, respectively.

These estimates give early indications of the constants to which the scale and shape parameters are close as well as the relationship between them. For instance, the estimated shape parameter (-0.526) is much closer to -1 and 0 than it is to 1. This, in turn, means the reciprocal Weibull distribution (which constrains the shape parameter to -1) and the log-normal distribution (which constrains the shape parameter to 0) are more plausible candidate distributions than the Weibull distribution (which constrains the shape parameter to 1).

With regard to the relationships between the scale and shape parameters, a model that constrains negative equality is $\delta = -q$ that seems to be more plausible compared to, for instance, a model that constrains reciprocal or negative reciprocal relationship. This is so because a reciprocal relationship would give a scale parameter of $1/(-0.526) = -1.90$, while a negative reciprocal relationship would yield $-(1/(-0.526)) = 1.90$ both of which are far from the freely estimated scale parameter 0.246. This, in turn, excludes models such as ammag and inverse ammag in favor of the inverse gamma model.

The above closeness of the special case models to the more general EGG model is also reflected in the values of log predictive density scores (LPDS) given in the last columns of each model. For instance, the LPDS of the EGG model is -4584, while that of the closest model (the inverse gamma) is -4594. On the other hand, the LPDS for ammag and inverse ammag are -5705 and -5184, respectively, which are far from that of the EGG.

Table 2 Posterior means (and 95 % hpd) of estimated effects in the selected models

		Women (EGG model)			Men (Inverse gamma)			Combined (EGG model)		
Covariate	Levels	Estimate	2.5 %	97.5 %	Estimate	2.5 %	97.5 %	Estimate	2.5 %	97.5 %
	No Educ	0 (ref.)	–	–	0 (ref.)	–	–	0 (ref.)	–	–
	Primary	0.015	0.002	0.028	−0.016	−0.04	0.009	0.042	0.029	0.055
Education	Middle	0.138	0.123	0.153	0.023	−0.001	0.047	0.186	0.007	0.2
	Secon	0.251	0.234	0.269	0.097	0.071	0.122	0.315	0.008	0.331
	PostSec	0.425	0.39	0.461	0.138	0.105	0.171	0.518	0.493	0.466
	Capital	0 (ref.)	–	–	0 (ref.)	–	–	0 (ref.)	–	–
Residence	Other	−0.086	−0.069	−0.052	−0.061	d	−0.035	−0.052	−0.068	−0.036
	Rural	−0.079	−0.094	−0.062	−0.137	−0.161	−0.111	−0.042	−0.058	−0.026

Table 3 Posterior means (and 95 % hpd) of covariate effects on log time to event: Women

	Educational Level (ref: No Educ)				Residence (ref: Capital)		Scale & Shape		GoF
Model	Primary	Middle	Secon	PostSec	Other Towns	Rural	Scale (δ)	Shape (q)	LPDS
EGG	0.015	0.138	0.251	0.425	−0.069	−0.079	0.246	−0.526	−4584
	(0.002, 0.028)	(0.123, 0.153)	(0.234, 0.269)	(0.39, 0.461)	(−0.086, −0.052)	(−0.094, −0.062)	(0.242, 0.25)	(−0.574, −0.476)	
Exponential	0.121	0.526	0.774	1.187	−0.126	−0.112	1	1	−6201
	(0.062, 0.179)	(0.454, 0.599)	(0.691,0.853)	(0.984, 1.398)	(−0.202, −0.040)	(−0.189, −0.031)	1	1	
Inv. Exp	0.066	0.321	0.488	0.755	−0.087	−0.079	1	−1	−5985
	(0.013, 0.119)	(0.264, 0.38)	(0.425, 0.55)	(0.623, 0.895)	(−0.155, −0.021)	(−0.141, −0.015)			
Weibull	**0.007**	0.116	0.228	0.42	−0.146	−0.184	0.267	1	−4892
	(−0.009, **0.022**)	(0.096, 0.136)	(0.206, 0.25)	(0.364, 0.478)	(−0.168, −0.124)	(−0.206, −0.162)	(0.263, 0.271)	1	
Rec. Weibull	0.021	0.146	0.247	0.419	−0.054	−0.053	0.243	−1	−4614
	(0.008, 0.032)	(0.132, 0.16)	(0.231, 0.263)	(0.382, 0.452)	(−0.07, −0.037)	(−0.069, −0.038)	(0.239, 0.247)		
Rayleigh	0.045	0.258	0.416	0.673	−0.129	−0.145	0.5	1	−5315
	(0.016, 0.075)	(0.222, 0.295)	(0.374, 0.459)	(0.57, 0.778)	(−0.171, −0.088)	(−0.186, −0.106)			
Inv. Rayleigh	0.032	0.194	0.322	0.524	−0.071	−0.073	0.5	−1	−5101
	(0.005, 0.06)	(0.165, 0.222)	(0.291, 0.353)	(0.453, 0.593)	(−0.102, −0.039)	(−0.105, −0.041)			
Stand. Gamma	0.008	0.125	0.242	0.414	−0.109	−0.134	q	**0.252**	−4665
	(−0.007, 0.022)	(0.109, 0.142)	(0.222, 0.262)	(0.371, 0.458)	(−0.129, −0.09)	(−0.152, −0.114)		**(0.248, 0.256)**	
Gamma	0.081	0.371	0.555	0.848	−0.094	−0.087	$1, q \neq 0$	−0.388	−5939
	(0.027, 0.135)	(0.311, 0.429)	(0.489, 0.622)	(0.705, 0.995)	(−0.162, −0.029)	(−0.155, −0.02)		(−0.444, −0.33)	
Inv. Gamma	**0.014**	0.135	0.25	0.423	−0.081	−0.097	$-q$	−0.251	−4594
	(0, 0.027)	(0.119, 0.152)	(0.233, 0.267)	(0.383, 0.463)	(−0.099, −0.063)	(−0.114, −0.079)		(−0.255, −0.247)	
Ammag	0.049	0.268	0.448	0.808	−0.175	−0.206	$1/q$	**2.361**	−5705
	(0.021, 0.076)	(0.225, 0.314)	(0.395, 0.502)	(0.66, 0.972)	(−0.221, −0.127)	(−0.25, −0.158)		**(2.31, 2.416)**	
Inv. Ammag	0.031	0.165	0.221	0.373	−0.045	−0.016	$-1/q$	−2.986	−5184
	(0.013, 0.05)	(0.145, 0.185)	(0.199, 0.243)	(0.326, 0.415)	(−0.065, −0.023)	(−0.037, 0.005)		(−3.062, −2.921)	
Log normal	**0.012**	0.131	0.248	0.421	−0.094	−0.114	0.252	0	−4621
	(−0.003, **0.026**)	(0.114, 0.146)	(0.229, 0.266)	(0.381, 0.461)	(−0.112, −0.077)	(−0.133, −0.097)	(0.249, 0.257)		
Half-normal	0.087	0.415	0.63	1.006	−0.136	−0.139	$\sqrt{2}/2$	$\sqrt{2}$	−5898
	(0.044, 0.13)	(0.358, 0.475)	(0.563, 0.694)	(0.836, 1.185)	(−0.199, −0.07)	(−0.2, −0.075)			

Table 4 Posterior means (and 95 % hpd) of covariate effects on log time to event: Men

	Educational Level (ref: No Educ)				Residence (ref: Capital)		Scale & Shape		GoF
Model	Primary	Middle	Secon	PostSec	Other Towns	Rural	Scale (δ)	Shape (q)	LPDS
EGG	−0.018	**0.022**	0.096	0.138	−0.062	−0.137	0.235	−0.199	−1754
	(−0.042, 0.005)	(−0.002, 0.049)	(0.069, 0.122)	(0.106, 0.171)	(−0.089, −0.035)	(−0.164, −0.11)	(0.228, 0.241)	(−0.289, −0.106)	
Exponential	0.123	0.604	0.861	0.558	−0.099	**−0.135**	1	1	−2441
	(0.014, 0.229)	(0.49, 0.722)	(0.726, 0.992)	(0.389, 0.72)	(−0.238, 0.038)	**(−0.272, 0)**			
Inv. Exp	**0.074**	0.307	0.44	0.344	−0.078	−0.126	1	−1	−2303
	(−0.025, 0.168)	(0.212, 0.396)	(0.345, 0.535)	(0.212, 0.475)	(−0.18, 0.019)	(−0.22, −0.022)			
Weibull	**−0.062**	**−0.037**	0.037	0.066	−0.055	−0.156	0.22	1	−1829
	(−0.085, −0.038)	**(−0.061, −0.011)**	(0.006, 0.067)	(0.027, 0.106)	(−0.086, −0.023)	(−0.191, −0.126)	(0.214, 0.225)		
Rec. Weibull	**−0.001**	0.037	0.108	0.153	−0.038	−0.11	0.241	−1	−1782
	(−0.024, 0.024)	(0.012, 0.06)	(0.083, 0.132)	(0.123, 0.184)	(−0.064, −0.014)	(−0.135, −0.086)	(0.235, 0.247)		
Rayleigh	**0.01**	0.193	0.341	0.265	−0.074	−0.142	0.5	1	−2083
	(−0.045, 0.06)	(0.133, 0.254)	(−0.14, −0.005)	(−0.209, −0.077)	(0.272, 0.405)	(0.18, 0.349)			
Inv. Rayleigh	0.02	0.115	0.206	0.215	−0.057	−0.117	0.5	−1	−1970
	(−0.029, 0.07)	(0.065, 0.162)	(0.157, 0.255)	(0.15, 0.277)	(−0.108, −0.004)	(−0.167, −0.069)			
Stand. Gamma	**−0.031**	**0.008**	0.081	0.12	−0.066	−0.15	q	**0.228**	−1763
	(−0.054, −0.008)	**(−0.017, 0.031)**	(0.054, 0.107)	(0.086, 0.153)	(−0.096, −0.041)	(−0.177, −0.123)		**(0.223, 0.234)**	
Gamma	**0.084**	0.344	0.492	0.376	−0.081	−0.128	$1, q \neq 0$	−0.615	−2297
	(−0.013, 0.177)	(0.252, 0.432)	(0.391, 0.594)	(0.241, 0.514)	(−0.184, 0.015)	(−0.227, −0.034)		(−0.707, −0.527)	
Inv. Gamma	**−0.016**	**0.023**	0.097	0.138	−0.061	−0.137	$-q$	−0.235	−1754
	(−0.04, 0.009)	**(−0.001, 0.047)**	(0.071, 0.122)	(0.105, 0.171)	(−0.087, −0.035)	(−0.161, −0.111)		(−0.241, −0.229)	
Ammag	**−0.042**	0.179	0.438	0.214	−0.026	−0.142	$1/q$	**2.86**	−2253
	(−0.084, 0)	(0.107, 0.253)	(0.351, 0.524)	(0.115, 0.313)	(−0.122, 0.068)	(−0.237, −0.049)		**(2.7, 3.02)**	
Inv. Ammag	**0.008**	0.049	0.113	0.147	−0.002	−0.075	$-1/q$	**−2.923**	−1998
	(−0.024, 0.039)	(0.017, 0.08)	(0.082, 0.141)	(0.104, 0.189)	(−0.035, 0.031)	(−0.106, −0.046)		**(−3.033, −2.815)**	
Log normal	**−0.023**	**0.016**	0.091	0.131	−0.065	−0.143	0.233	0	−1756
	(−0.049, 0.002)	**(−0.009, 0.04)**	(0.065, 0.117)	(0.097, 0.165)	(−0.093, −0.037)	(−0.171, −0.117)	(0.227, 0.24)		
Half-normal	0.064	0.435	0.673	0.436	−0.084	−0.139	$\sqrt{2}/2$	$\sqrt{2}$	−2335
	(−0.012, 0.146)	(0.343, 0.527)	(0.573, 0.778)	(0.309, 0.561)	(−0.192, 0.031)	(−0.253, −0.028)			

Table 5 Posterior means (and 95 % hpd) of covariate effects on log time to event: combined sample

	Educational Level (ref: No Educ)				Residence (ref: Capital)		Scale & Shape		GoF
Model	Primary	Middle	Secon	PostSec	Other Towns	Rural	Scale (δ)	Shape (q)	LPDS
EGG	0.042	0.186	0.315	0.493	−0.052	−0.042	0.285	−0.525	−6695
	(0.029, 0.055)	(0.172, 0.2)	(0.299, 0.331)	(0.466, 0.518)	(−0.068, −0.036)	(−0.058, −0.026)	(0.281, 0.289)	(−0.575, −0.475)	
Exponential	0.162	0.643	0.936	1.021	−0.078	**−0.03**	1	1	−8682
	(0.112, 0.21)	(0.582, 0.702)	(0.865,1.008)	(0.905, 1.141)	(−0.146, −0.009)	**(−0.097, 0.04)**			
Inv. Exp	0.102	0.403	0.591	0.742	−0.056	**−0.024**	1	−1	−8331
	(0.057, 0.151)	(0.356, 0.449)	(0.544, 0.642)	(0.658, 0.828)	(−0.111, −0.005)	**(−0.076, 0.027)**			
Weibull	0.02	0.147	0.278	0.412	−0.077	−0.102	0.294	1	−7068
	(0.005, 0.035)	(0.128, 0.165)	(0.257, 0.298)	(0.371, 0.451)	(−0.098, −0.057)	(−0.122, −0.082)	(0.290, 0.298)		
Rec. Weibull	0.042	0.189	0.308	0.491	−0.039	−0.025	0.279	−1	−6729
	(0.029, 0.055)	(0.175, 0.202)	(0.293, 0.322)	(0.468, 0.518)	(−0.055, −0.025)	(−0.039, −0.01)	(0.275, 0.283)		
Rayleigh	0.068	0.305	0.488	0.611	−0.078	−0.073	0.5	1	−7508
	(0.041, 0.094)	(0.275, 0.338)	(0.454, 0.524)	(0.55, 0.673)	(−0.113, −0.041)	(−0.108, −0.038)			
Inv. Rayleigh	0.058	0.246	0.389	0.563	−0.047	−0.03	0.5	−1	−7190
	(0.036, 0.082)	(0.22, 0.27)	(0.362, 0.413)	(0.519, 0.605)	(−0.075, −0.021)	(−0.056, −0.004)			
Stand. Gamma	0.035	0.173	0.307	0.464	−0.07	−0.075	q	0.288	−6801
	(0.021, 0.049)	(0.157, 0.189)	(0.29, 0.324)	(0.433, 0.495)	(−0.09, −0.053)	(−0.094, −0.057)		(0.285, 0.292)	
Gamma	0.119	0.452	0.66	0.8	−0.062	**−0.027**	$1, q \neq 0$	−0.454	−8280
	(0.075, 0.164)	(0.405, 0.497)	(0.608, 0.715)	(0.712, 0.891)	(−0.113, −0.008)	**(−0.077, 0.025)**		(−0.499, −0.407)	
Inv. Gamma	0.041	0.184	0.318	0.491	−0.057	−0.05	$-q$	−0.289	−6703
	(0.027, 0.055)	(0.17, 0.199)	(0.301, 0.333)	(0.464, 0.518)	(−0.074, −0.04)	(−0.067, −0.034)		(−0.293, −0.285)	
Ammag	0.035	0.288	0.517	0.595	−0.07	−0.108	$1/q$	2.304	−8074
	(0.01, 0.062)	(0.251, 0.326)	(0.472, 0.563)	(0.515, 0.682)	(−0.113, −0.027)	(−0.15, −0.069)		(2.256, 2.353)	
Inv. Ammag	0.046	0.2	0.286	0.46	−0.034	**0.003**	$-1/q$	−2.738	−7346
	(0.029, 0.063)	(0.182, 0.218)	(0.268, 0.303)	(0.428, 0.49)	(−0.052, −0.016)	**(−0.015, 0.021)**		(−2.785, −2.688)	
Log normal	0.039	0.181	0.316	0.48	−0.065	−0.063	0.291	0	−6738
	(0.025, 0.054)	(0.166, 0.197)	(0.298, 0.333)	(0.451, 0.508)	(−0.083, −0.048)	(−0.079, −0.046)	(0.287, 0.295)		
Half-normal	0.114	0.495	0.754	0.841	−0.08	−0.057	$\sqrt{2}/2$	$\sqrt{2}$	−8294
	(0.079, 0.153)	(0.446, 0.542)	(0.7, 0.808)	(0.743, 0.936)	(−0.138, −0.027)	(−0.118, −0.006)			

Another point worth noting is that the estimates of the covariate effects and their associated 95% hpd are much alike in the models that are close to each other (in terms of estimated scale and/or shape parameters or in terms of LPDS) than those estimates that are far apart.

Thus, for the women data, it would not make much difference if we base our conclusions on the estimates from the EGG model or the inverse gamma model though a formal test would favor the larger EGG model.

The results for men shown in Table 4 can be interpreted similarly. Here, the scale and shape parameters estimated freely from the data in the EGG model are $\delta = 0.235$ and $q = -0.199$, respectively. Again, the inverse gamma model that imposes a negative relationship between the scale and shape parameters ($\delta = -q$) seems to be much more plausible than any other model. In fact, a closer look at the LPDS values shows that it even outperforms the larger EGG model though the difference in LPDS is marginal.

Hence, for the men data, we have a very strong evidence to base our conclusions on the results from the inverse gamma model that, of course, are identical to those from the EGG model.

Last, the results for the combined sample are shown in Table 5. Similar reasoning as in the above leads to the choice of EGG model or the inverse gamma model though a formal test would favor the larger EGG model. That the results for the combined sample reflected those for women are not surprising because women constitute about two-third of the combined sample.

The final estimates of covariate effects and their associated 95% hpd from our chosen models for respective data sets are summarized in Table 2.

The results in Table 2 show that there are significant differentials in entry to first marriage across women's educational level and residence where lower education and rural residence are associated with higher intensities of marriage. For men, the educational differences are less pronounced as there is no significant difference in the intensities of entry to first marriage between those with no education and those with primary or middle education. The residential differential is, however, still significant. The results for the combined sample follow those of women because, as mentioned before, women constitute majority in the combined sample.

5 Summary and Concluding Remarks

In this chapter, we presented the extended generalized gamma (EGG) model for survival data with censored observations. Previous works have shown that five known models can be treated as special cases of the EGG model by constraining the scale parameter, shape parameter, or both to some constants. In the present chapter, we extended the EGG model further to include 13 special case models. This was achieved by imposing relationships between the scale and shape parameters in addition to constraining them to some constants.

The issues were illustrated with data on entry into first marriage among Eritrean men and women based on data from the 2010 Eritrean Population and Health Survey

(EPHS 2010). Inference was fully Bayesian using a random walk Metropolis–Hastings algorithm to sample from the posterior distribution, and we compared the models with each other and relative to the more general EGG model using the log predictive density score (LPDS).

The application demonstrates that the further extended family of distributions provides a wide range of alternatives for a baseline distribution in the analysis of survival data with censored observations. For instance, we found that the inverse gamma model, where we impose the scale parameter to be the negative of the shape parameters ($\delta = -q$), fits the men data best and outperforms the EGG model. It also performs well in the women data and the combined sample though the evidence is not as strong as in the men data. This was in accordance with the freely estimated values of the scale and shape parameters in the EGG model.

The empirical results in the final selected models reveal significant differentials in the pace of entry to first marriage across women's educational levels and residence. As would be expected, lower education and rural residence is associated with higher intensities of marriage. Educational differentials are, however, less pronounced for men as there was no significant difference in the intensities of entry to first marriage between those with no education (the baseline group) and those with primary or middle education. The residential differential was still significant in the men's data. When we analyzed the combined data, the results followed those of women due, mainly, to the fact that women constitute about two-third in the combined sample.

It may be worth noting that the educational level of the individuals refers to what is achieved by the survey time. As such, it is anticipatory in the sense that the reported educational level might have been achieved after the event of interest. But, our aim here is to demonstrate the models and methods empirically, and the anticipatory nature of education does not affect our purpose. Ghilagaber and Koskinen (2009), Ghilagaber and Larsson (2019), and Munezero and Ghilagaber (2022a) study potential biases due to the use of anticipatory covariates and how to account for that.

Our analysis was based on the tacit assumption that the survivor function $S(t)$ tends to 0 as the study period gets longer. This, in turn, means that we have assumed all individuals will experience the event of interest sooner or later. This may not be true for the event in our illustrative example (marriage) as there may be some individuals who may never marry for various reasons. Future works may, therefore, consider accounting for such long-term survivors (those who may never experience the event of interest). This can be achieved by using, for instance, a mixture model consisting of a hazard/intensity model for those who experienced the event or may experience it in the future and a logistic model for the probability of being long-term survivor (never experiencing the event).

Appendix 1: Density Functions, $f(t)$, and Survival Functions, $S(t)$, of Special Cases in the Extended Generalized Gamma Model

Distribution	$f(t)$	$S(t)$
Exponential	$e^{-\mu}e^{-e^{-\mu}t}$	$e^{-e^{-\mu}t}$
I. Exponential	$e^{\mu}e^{-\frac{e^{\mu}}{t}}/t^2$	$1-e^{-\frac{e^{\mu}}{t}}$
Weibull	$\frac{1/\delta}{e^{\mu}}\left(\frac{t}{e^{\mu}}\right)^{1/\delta-1}e^{-\left(\frac{t}{e^{\mu}}\right)^{1/\delta}}$	$e^{-t^{1/\delta}e^{-\mu/\delta}}$
R. Weibull	$\frac{1/\delta}{e^{\mu}}\left(\frac{t}{e^{\mu}}\right)^{-1/\delta-1}e^{-\left(\frac{t}{e^{\mu}}\right)^{-1/\delta}}$	$1-e^{-t^{-1/\delta}e^{\mu/\delta}}$
Rayleigh	$\frac{2t}{e^{2\mu}}e^{-\frac{t^2}{e^{2\mu}}}$	$e^{-\frac{t^2}{e^{2\mu}}}$
I. Rayleigh	$\frac{2e^{2\mu}}{t^3}e^{-\frac{e^{2\mu}}{t^2}}$	$1-e^{-\frac{e^{2\mu}}{t^2}}$
S. Gamma	$1/\Gamma(q^{-2})\left(q^{-2}e^{-\mu}\right)^{q^{-2}}t^{q^{-2}-1}e^{-q^{-2}e^{-\mu}t}$ [a]	$1-\gamma\left(q^{-2},tq^{-2}e^{-\mu}\right)/\Gamma(q^{-2})$ [b]
Gamma	$(q>0)\,qt^{q-1}\frac{1}{\Gamma(q^{-2})}\left(q^{-2}e^{-q\mu}\right)^{q^{-2}}(t^q)^{q^{-2}-1}e^{-q^{-2}e^{-q\mu}t^q}$	$1-\gamma\left(q^{-2},t^qq^{-2}e^{-q\mu}\right)/\Gamma(q^{-2})$
	$(q<0)\,-qt^{q-1}\frac{1}{\Gamma(q^{-2})}\left(q^{-2}e^{-q\mu}\right)^{q^{-2}}(t^q)^{q^{-2}-1}e^{-q^{-2}e^{-q\mu}t^q}$	$\gamma\left(q^{-2},t^qq^{-2}e^{-q\mu}\right)/\Gamma(q^{-2})$
I. Gamma	$t^{-2}\frac{1}{\Gamma(q^{-2})}\left(q^{-2}e^{\mu}\right)^{q^{-2}}\left(t^{-1}\right)^{q^{-2}-1}e^{-q^{-2}e^{\mu}t^{-1}}$	$\gamma\left(q^{-2},t^{-1}q^{-2}e^{\mu}\right)/\Gamma(q^{-2})$
Ammag	$q^2t^{q^2-1}\frac{1}{\Gamma(q^{-2})}\left(q^{-2}e^{-q^2\mu}\right)^{q^{-2}}\left(t^{q^2}\right)^{q^{-2}-1}e^{-q^{-2}e^{-q^2\mu}t^{q^2}}$	$1-\gamma\left(q^{-2},\frac{t^{q^2}}{q^2}e^{-q^2\mu}\right)/\Gamma(q^{-2})$
I. Ammag	$q^2t^{-q^2-1}\frac{1}{\Gamma(q^{-2})}\left(q^{-2}e^{q^2\mu}\right)^{q^{-2}}\left(t^{-q^2}\right)^{q^{-2}-1}e^{-q^{-2}e^{q^2\mu}t^{-q^2}}$	$\gamma\left(q^{-2},\frac{t^{-q^2}}{q^2}e^{q^2\mu}\right)/\Gamma(q^{-2})$
Log normal	$\frac{1}{t\delta\sqrt{2\pi}}e^{-\frac{(logt-\mu)^2}{2\delta^2}}$	$\frac{1}{2}(1+erf(\frac{x-\mu}{\sqrt{2}\delta}))$
Half-normal	$\frac{1}{e^{\mu}}\sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2e^{2\mu}}}$	$1-erf(\frac{x}{\sqrt{2}e^{\mu}})$ [c]

[a] $\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}dt$

[b] $\gamma(s,r)=\int_0^{r}x^{s-1}e^{-x}dx$

[c] $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^{x}e^{-z^2}dz$

Appendix 2: Proof of Lemma 1

Proof:

$$E(T^r) = \int_0^\infty t^r \frac{|q|}{\delta} t^{\frac{q}{\delta}-1} \frac{[q^{-2}(e^{-\mu})^{\frac{q}{\delta}}]^{q^{-2}}}{\Gamma(q^{-2})} (t^{\frac{q}{\delta}})^{q^{-2}-1} e^{-q^{-2}(e^{-\mu})^{\frac{q}{\delta}} t^{\frac{q}{\delta}}} dt.$$

Let $x = q^{-2}(e^{-\mu})^{\frac{q}{\delta}} t^{\frac{q}{\delta}}$; then $t = (q^2 e^{\frac{q\mu}{\delta}} x)^{\frac{\delta}{q}}$,

$$E(T^r) = \frac{[(q^2 e^{\frac{q\mu}{\delta}})^{\frac{\delta}{q}}]^r}{\Gamma(q^{-2})} \int_0^\infty x^{\frac{r\delta}{q}+q^{-2}-1} e^{-x} dx$$

$$= \frac{[(q^2 e^{\frac{q\mu}{\delta}})^{\frac{\delta}{q}}]^r}{\Gamma(q^{-2})} \Gamma(q^{-2} + \frac{r\delta}{q}) \quad (q^{-2} + \frac{r\delta}{q} > 0)$$

When $r = 1$, we get $E(T)$, and when $r = 2$, we get $E(T^2)$. Using these, we have $V(T) = E(T^2) - E^2(T)$. ■

Appendix 3: R Program Codes for Bayesian Inference

```
#################################################################
############Random Walk Metropolis algorithm####################
#################################################################
library(MASS)
library(mvtnorm)

#Calculate log-likelihood and the Hessian evaluated at the mode

calculate.loglike <- function(beta){

  beta1 <- beta[1]+beta[2]*Prim+beta[3]*Middle+beta[4]*Secon+
           beta[5]*PostSec+beta[6]*Other+beta[7]*Rural
  delta <- beta[8]
  q <- beta[9]

   error <- (log(y)-beta1)/delta
   s <- q^(-2)
   r <- exp(q*error)
   #Probability density function
   pdf <- function(r,s,q){
       if (q!=0) pdf <- abs(q)*r*dgamma(r,s,s)
       else pdf <- dnorm(error,0,1)
   pdf
   }
   #Survival function
```

```
   surv <- function(r,s,q){
        if (q!=0) surv <- (q>0)*(1-pgamma(r,s,s))+
                          (q<0)*(pgamma(r,s,s))
        else surv <- 1-pnorm(error,0,1)
  surv
  }
   #log-likelihood function
   loglike <- function(r,s,q){
    loglike <- sum(d*(log(pdf(r,s,q))-log(delta))+
               (1-d)*log(surv(r,s,q)))
   loglike
  }

  list(loglike=loglike(r,s,q))
}

m=20000
beta0 <- c(5,0,0,0,0,0,0)
delta0 <- 1
q0 <- 0.12

theta0 <- c(beta0,delta0,q0)

mu.beta <- c(0,0,0,0,0,0,0) #prior
s.beta <- c(1000,1000,1000,1000,1000,1000,1000) #prior

theta <- matrix(nrow=m,ncol=9)
acc.prob_theta <- 0
current.theta <- theta0

for (t in 1:m){
  cur <- calculate.loglike(current.theta)

  prop.theta <- mvrnorm(1,current.theta,Sigma=vcov_mode)
  jump_1 <- -as.numeric(0.5*log(det(vcov_mode)))-
        0.5*t(prop.theta-current.theta)%*%solve(vcov_mode)
        %*%(prop.theta-current.theta)
  prop <- calculate.loglike(prop.theta)

  ll_1 <- prop$loglike
  jump_2 <- -as.numeric(0.5*log(det(vcov_mode)))-
        0.5*t(current.theta-prop.theta)%*%solve(vcov_mode)
        %*% (current.theta-prop.theta)
  ll_2 <- cur$loglike
  loga <- ll_1-ll_2+sum(dnorm(prop.theta[1:7],mu.beta,s.beta,
                        log=T))-
          sum(dnorm(current.theta[1:7],mu.beta,s.beta,log=T))+
          dgamma(prop.theta[8],1,1,log=T)-dgamma(current.theta[8],
                    1,1,log=T)+
          dnorm(prop.theta[9],0,1000,log=T)-dnorm(current.theta[9],
                  0,1000,log=T)+
          jump_2-jump_1

    u <- runif(1)
```

```
    u <- log(u)
     if (u < loga){
        current.theta <- prop.theta
        acc.prob_theta <- acc.prob_theta+1
       }
  theta[t,] <- current.theta
}

#################################################################
###########Log predictivedensity score (LPDS)#################
#################################################################

#5-fold cross validation
cv <- matrix(cbind(Age,1,Prim,Middle,Secon,PostSec,Other,Rural,
      Marr),ncol=9, nrow=length(Age))

index <- sample(length(y),size=length(y),replace=F)
index_1 <- index[1:2047]
index_2 <- index[2048:4095]
index_3 <- index[4096:6143]
index_4 <- index[6144:8191]
index_5 <- index[8192:10238]

new <- list(as.matrix(cv[index_1,]),as.matrix(cv[index_2,]),as.
            matrix(cv[index_3,]),
            as.matrix(cv[index_4,]),as.matrix(cv[index_5,]))

calculate.loglike <- function(b,delta,q,data){

   x<- data[,2:8]
   y <- data[,1]
   d <- data[,9]
   mu <- x%*%b

   #Probility density function
   pdf <- function(b,delta,q){
       if (q!=0) pdf <- abs(q)/delta*y^(q/delta-1)*
                        dgamma(y^(q/delta),q^(-2),q^(-2)*exp
                        (-mu)^(q/delta))
       else pdf <- dnorm(log(y),mu,delta^2)
   pdf
   }
   #Survival function
   surv <- function(b,delta,q){
       if (q!=0) surv <- (q>0)*(1-pgamma(y^(q/delta),q^(-2),q^(-2)
                         *exp(-mu)^(q/delta)))+
                         (q<0)*pgamma(y^(q/delta),q^(-2),q^(-2)
                         *exp(-mu)^(q/delta))
       else surv <- 1-pnorm(log(y),mu,delta^2)
   surv
   }
  loglike <- sum(d*(log(pdf(b,delta,q)))+(1-d)*log(surv(b,delta,q)))
  return(list(loglike=loglike))
 }
```

```
m<-20000

#initial values
beta0 <- mode[1:7]
delta0 <- mode[8]
q0 <- -0.15
theta0 <- c(beta0,delta0)

mu.beta <- c(0,0,0,0,0,0,0) #prior
s.beta <- c(1000,1000,1000,1000,1000,1000,1000) #prior

theta <- matrix(nrow=m,ncol=8)
q <- c()
pre <- c()
logpre <- c()
loglike <- c()
Cb <- c()
acc.prob_theta <- 0
acc.prob_q <- 0
current.theta <- theta0
current.q <- q0

   for  (k in 1:5){
   for (t in 1:m){
   prop.theta <- mvrnorm(1,current.theta,cov[1:8,1:8])
   ll_1 <- calculate.loglike(prop.theta[1:7],prop.theta[8],
           current.q,data=rbind(new[-k][[1]],new[-k][[2]],
           new[-k][[3]],new[-k][[4]]))$loglike
   ll_2 <- calculate.loglike(current.theta[1:7],current.theta[8],
           current.q,data=rbind(new[-k][[1]],new[-k][[2]],new[-k]
           [[3]],new[-k][[4]]))$loglike
   jump_1 <- -as.numeric(0.5*log(det(cov[1:8,1:8])))-
   0.5*t(prop.theta-current.theta)%*%solve(cov[1:8,1:8])%*%(prop.
   theta-current.theta)
   jump_2 <- -as.numeric(0.5*log(det(cov[1:8,1:8])))-
   0.5*t(current.theta-prop.theta)%*%solve(cov[1:8,1:8])
   %*%(current.theta-prop.theta)

   loga <- ll_1-ll_2+sum(dnorm(prop.theta[1:7],mu.beta,s.beta,
           log=T))-sum(dnorm(current.theta[1:7],mu.beta,s.beta,
           log=T))+dgamma(prop.theta[8],1,1,log=T)-dgamma(current.
           theta[8],1,1,log=T)+jump_2-jump_1

   u <- runif(1)
   u <- log(u)
     if (u < loga){
        current.theta <- prop.theta
        acc.prob_theta <- acc.prob_theta+1
       }
  theta[t,] <- current.theta

  prop.q <- rnorm(1,current.q,cov[9,9])
  ll_1 <- calculate.loglike(current.theta[1:7],current.theta[8],
```

```
          prop.q,data=rbind(new[-k][[1]],new[-k][[2]],new[-k][[3]],
          new[-k][[4]]))$loglike
  ll_2 <- calculate.loglike(current.theta[1:7],current.theta[8],
          current.q,data=rbind(new[-k][[1]],new[-k][[2]],new[-k]
          [[3]],new[-k][[4]]))$loglike
  loga <- ll_1-ll_2+dnorm(prop.q,0,1000,log=T)-dnorm(current.q,0,
          1000,log=T)
  u <- runif(1)
  u <- log(u)
  if (u < loga){
    current.q <- prop.q
    acc.prob_q <- acc.prob_q+1
  }
  q[t] <- current.q
  loglike[t] <- calculate.loglike(theta[t,1:7],theta[t,8],q[t],
                data=new[[k]])$loglike
  }
  Cb[k] <- mean(loglike)
  pre <- exp(loglike-Cb[k])
  burnin <- 5000
  logpre[k]<- Cb[k]+log(mean(pre[(burnin+1):m]))
}
lpds <- mean(logpre)
```

Acknowledgments Yuli Liang's research is partly supported by the strategic funding of Örebro University. Liang carried out part of the work while affiliated with Stockholm University. The data sets analyzed in this chapter were provided to the co-author (GG) by the National Statistics Office, Eritrea. He is grateful to its director Mr. Ainom Berhane and its senior statisticians Mr. Hagos Ahmed and Mr. Samuel Tesfamariam. The views expressed in the chapter are solely of the authors and do not express the views or opinions of the data source or its employees.

References

Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover publications.

Addison, J., & Portugal, P. (1987). On the distributional shape of unemployment duration. *The Review of Economics and Statistics, 69*(3), 520–526.

Berger, J. O., Bernardo, J. M., Sun, D. et al. (2015). Overall objective priors. *Bayesian Analysis, 10*, 189–221.

Bergström, R., & Edin, P. (1992). Time aggregation and the distributional shape of unemployment duration. *Journal of Applied Econometrics, 7*(1), 5–30.

Bergström, R., Engwall, L., & Wallerstedt, E. (1994). Organizational foundations and closures in a regulated environment: Swedish commercial banks 1831–1990. *Scandinavian Journal of Management, 10*(1), 29–48.

Bergström, R., Engwall, L., & Wallerstedt, E. (1997). The importance of flexible hazard functions in the analysis of organizational survival data–experiences from a cohort of Swedish commercial banks. *Quality and Quantity, 31*(1), 15–35.

Butler, R., & McDonald, J. (1986). Trends in unemployment duration data. *The Review of Economics and Statistics, 68*(4), 545–557.

Cox, C., Chu, H., Schneider, M., & Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine, 26*(23), 4352–4374.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–202.

Farewell, V., & Prentice, R. (1977). A study of distributional shape in life testing. *Technometrics, 19*(1), 69–75.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Ghilagaber, G. (2005). The extended generalized gamma model and its special cases: Applications to modeling marriage durations. *Quality and Quantity, 39*(1), 71–85.

Ghilagaber, G. (2018). Statistical analysis of educational and urban-rural differences in marriage-timing: Evidence from the 2010 Eritrean population and health survey for women and men. In *Proceedings of the International Conference on Eritrean Studies, 20–22 July 2016* (pp. 1097–1124).

Ghilagaber, G., Elisa, W., & Gyimah, S. O. (2014). A family of flexible parametric duration functions and their applications to modeling child-spacing in sub-Saharan Africa. In Advanced Techniques for Modelling Maternal and Child Health in Africa. Springer Series on Demographic Methods and Population Analysis (Vol. 34, pp. 185–209).

Ghilagaber, G., & Koskinen, J. H. (2009). Bayesian adjustment of anticipatory covariates in analyzing retrospective data. *Mathematical Population Studies, 16*(2), 105–130.

Ghilagaber, G., & Larsson, R. (2019). Maximum likelihood adjustment of anticipatory covariates in the analysis of retrospective data. Research report 2019–1. Stockholm: Department of Statistics, Stockholm University.

Godha, D., Gage, A. J., Hotchkiss, D. R., & Cappa, C. (2016). Predicting maternal health care use by age at marriage in multiple countries. *Journal of Adolescent Health, 58*(5), 504–511.

Heleno, A., & Alberto, J. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian approach. *Statistics and Probability Letters, 4*(6), 325–332.

Li, F., Villani, M., & Kohn, R. (2010). Flexible modeling of conditional distributions using smooth mixtures of asymmetric student t densities. *Journal of Statistical Planning and Inference, 140*, 3638–3654.

Munezero, P., & Ghilagaber, G. (2022a). Dynamic Bayesian adjustment of anticipatory covariates in retrospective data: application to the effect of education on divorce risk. *Journal of Applied Statistics, 49*(6), 1382–1401.

Munezero, P., & Ghilagaber, G. (2022b). Dynamic Bayesian modelling of educational and residential differences in family initiation among Eritrean men and women. In *Modern Biostatistical Methods for Evidence-Based Global Health Research*. Springer Series on Emerging Topics in Statistics and Biostatistics.

National-Statistics-Office-Eritrea & Fafo-AIS (2013). *Eritrea Population and Health Survey 2010*. New York: National Statistics Office and Fafo Institute for Applied International Studies.

National-Statistics-Office-Eritrea & Macro-International-Inc. (1997). *Eritrea Demographic and Health Survey 1995*. New York: National Statistics Office and Macro International.

National-Statistics-Office-Eritrea & Macro-International-Inc. (2003). *Eritrea Demographic and Health Survey 2002*. New York: National Statistics and Evaluation Office and ORC Macro.

Peng, Y., Dear, K., & Denham, J. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine, 17*(8), 813–830.

Prentice, R. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika, 61*(3), 539.

Raj, A. (2010). When the mother is a child: the impact of child marriage on the health and human rights of girls. *Archives of Disease in Childhood, 95*(11), 931–935.

Ramos, P. L., Achcar, J. A., Moala, F. A., Ramos, E., & Louzada, F. (2017). Bayesian analysis of the generalized gamma distribution using non-informative priors. *Statistics, 51*, 824–843.

Stacy, E. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics, 33*(3), 1187–1192.

Tsionas, E. (2001). Exact inference in four-parameter generalized gamma distributions. *Communications in Statistics-Theory and Methods, 30*(4), 747–756.

Van Noortwijk, J. (2001). Bayes estimates of flood quantiles using the generalised gamma distribution. In *System and Bayesian Reliability; Essays in Honor of Professor Richard E. Barlow on his 70th Birthday* (pp. 351–374).

Villani, M., Kohn, R., & Giordani, P. (2009). Regression density estimation using smooth adaptive Gaussian mixtures. *Journal of Econometrics, 153*, 155–173.

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine, 11*(14–15), 1871–1879.

Dynamic Bayesian Modeling of Educational and Residential Differences in Family Initiation Among Eritrean Men and Women

Parfait Munezero and Gebrenegus Ghilagaber

Abstract We propose a dynamic Bayesian survival model for analyzing differentials in the timing of family initiation. Such formulation relaxes the strong assumption of constant hazard ratio in conventional proportional hazards models and allows covariate effects to vary over time. Inference is fully Bayesian, and efficient sequential Monte Carlo (particle filter) is used to sample from the posterior distribution. We illustrate the proposed model with data on entry into first marriage among Eritrean men and women surveyed in the 2010 Eritrean Population and Health Survey. Results from the conventional proportional hazards model indicate significant differences in family initiation among all educational and residential groups. In the dynamic model, on the other hand, only one educational and one residential group among the women and only one residential group among the men differ from their respective baseline groups. Since the empirical relative intensities of entry into first marriage vary across respondents' ages, we argue that the proposed dynamic model captures differentials in family initiation more accurately.

Keywords Survival models · Time to event data · Censoring · Proportional hazards models · Semi parametric survival models · Hazard/Intensity function · Survival function · Intensity ratios · Exposure time · Bayesian inference · Dynamic modelling · Log normal priors · Inverse gamma priors · Wishart priors · Discount Factor · Generalized additive models · Polynomial splines · Non linear models · Markov chain Monte Carlo (MCMC) · Sequential Monte Carlo (SMS) · Particle Filter (PF) · Random walk · Prior distribution · Likelihood · Posterior distribution · Piece wise exponential model · Eritrea · Entry into marriage · Nuptiality · Demographic and Health Surveys (DHS) · Observational studies · Retrospective Surveys · Educational gradients in family initiation · Residential gradients in family initiation

P. Munezero · G. Ghilagaber (✉)
Stockholm University, Department of Statistics, Stockholm, Sweden
e-mail: Parfait.Munezero@ericsson.com; Gebre@stat.su.se

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_12

1 Introduction

Whether differences in survival functions are constant or change over time is an issue of interest in the analysis of time-to-event data. For instance, tests for equality between survival functions (typically the log-rank and generalized Wilcoxon tests) are based on the difference between the observed and expected number of events at each event time. But, they may yield different results (and lead to different conclusions) depending on whether the differences between the empirical survival curves are uniform or vary across the observation period. This is so because the generalized Wilcoxon test is a weighted version of the log-rank test where the differences between the observed and expected events are weighted by the "risk set" (the number of individuals at "risk" of experiencing the event) at each time. The risk set is large at the beginning of the study but is depleted over time as individuals experience the event. Thus, differences between survival curves at the beginning of the study are weighted heavier than differences at the end of the study. If the empirical survival curves are parallel (their differences are uniform over the study period), the two tests lead to the same conclusions. Else, one test statistic may be larger or smaller than the other depending on the extent to which the survival curves differ and where in the observation period they differ.

Similarly, the conventional proportional hazards model (Cox, 1972), assumes that the ratio of hazards of any two study groups is constant over the study period. In other words, the model assumes that the hazards of the study groups are parallel over the study period. This is clear in the model specification where the hazard (intensity) function is expressed as a product of a baseline hazard, $\lambda_0(t)$, which is a function of time, t, and a relative hazard that is a function of a vector of covariates, $\mathbf{x}$:

$$\lambda\left(t|\,\mathbf{x}\right) = \lambda_0\left(t\right)\exp\left(\mathbf{x}'\boldsymbol{\beta}\right), \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of parameters to be estimated.

In real-life situations, however, the effect of a covariate (and, hence, the ratio of the hazards) may change over time, especially when the observation period is long. For instance, a treatment may have immediate effect on a patient's survival chances, but its beneficial effect may diminish with time. Similarly, education may have deferring effect on entry into marriage at younger ages, but such educational difference may disappear at older ages when those who eventually marry had to do it sooner or later.

In order to accommodate such real-life situations, we need to relax the restrictive assumption in proportional hazards models given in Eq. (1). Thus, the above model can be extended to a more general hazard model where the parameters are allowed to vary over time:

$$\lambda\left(t|\mathbf{x}\right) = \lambda_0(t)\exp\left(\mathbf{x}'\boldsymbol{\beta}(t)\right). \tag{2}$$

This model is a special case of the generalized additive models described in Fahrmeir and Kneib (2011) and Hennerfeind et al. (2006) where nonlinear effects of continuous covariates, spatial effects, and frailty terms are incorporated.

An important issue that needs to be addressed in the extended model given in Eq. (2) is how to express $\beta(t)$. A general approach, such as that described in Hennerfeind et al. (2006) and van Houwelingen and Putter (2011), expresses $\beta(t)$ using polynomial splines. Another common and simpler approach, which is a special case of the spline model, defines $\beta(t)$ using piece-wise constant functions. This results in the semi-parametric piece-wise exponential model described in Gamerman (1991). The study period is partitioned into many small intervals, and $\beta(t)$ is assumed to be constant within each interval but may vary between intervals. This approach has been applied to analyze continuous survival times in Gamerman (1991), Hemming and Shaw (2002, 2005) and Wagner (2011) and discrete survival times in Fahrmeir (1994) and Fahrmeir and Wagenpfeil (1996).

The evolution of $\beta(t)$ across the intervals partitioning the study period is modeled using a random walk prior process. This process allows the effect parameters to adapt to any changes in the hazard function that may occur along time. On the other hand, it induces non-linearity and correlations among parameters and makes the posterior analytically intractable. To address these issues, we use sequential Monte Carlo (SMC), also known as particle filter in the literature, to sample from our posterior. As described in Munezero (2022), the particle filter is more suitable for nonlinear dynamic models. Further, Doucet et al. (2001) argue that it does not require any transformation of the model and also has computational advantage compared to conventional Markov Chain Monte Carlo (MCMC) methods.

The rest of the chapter is organized as follows. We introduce our illustrative data sets in Sect. 2 and perform analyses using conventional proportional hazards models for later comparison with dynamic models. In Sect. 3, we present the dynamic survival model in the Bayesian framework and describe the likelihood, prior, and the resulting posterior distributions. In Section 4, we fit the dynamic model to our data sets, discuss the results, and compare them with those obtained from standard models. We summarize our findings in Sect. 5 by way of concluding remarks and suggestions for future works in the area.

2 The Data Set and Preliminary Analyses with Standard Models

Age at marriage and child spacing are of substantive interest because of their well-documented association with women's health directly (Raj, 2010), or indirectly through their negative impact on health care utilization (Godha et al., 2016).

The data used for illustration in this chapter come from the 2010 Eritrea Population and Health Survey, EPHS2010, described in National-Statistics-Office-Eritrea and Fafo-AIS (2013). The EPHS2010 was designed as a follow-up to its

Table 1 Summary statistics of the data sets

		Women			Men			Combined sample		
Covariate	Levels	n	Events	% Event	n	Event	% Event	n	Event	% Event
Education	No Educ	4186	3799	90.75	1050	892	84.95	5236	4691	89.59
	Primary	2055	1634	79.51	803	543	67.62	2858	2177	76.17
	Middle	1827	1006	55.06	1209	455	37.63	3036	1461	48.12
	Sec	1894	886	46.78	1516	461	30.41	3410	1347	39.50
	PostSec	276	96	34.78	442	218	49.32	718	314	43.73
Residence	Capital	1819	969	53.27	931	344	36.95	2750	1313	47.75
	Other Towns	2504	1739	69.45	1257	565	44.95	3761	2304	61.26
	Rural Areas	5915	4713	79.68	2833	1660	58.60	8748	6373	72.85
	Total	10,238	7421	72.49	5021	2569	51.17	15,259	9990	65.47

predecessors—the 1995 and 2002 Demographic and Health Surveys described in National-Statistics-Office-Eritrea and Macro-International-Inc. (1997) and National-Statistics-Office-Eritrea and Macro-International-Inc. (2003), respectively, and updates the information from the previous surveys as well as provides findings on some new topics of interest.

The EPHS2010 was conducted between January and July 2010 and surveyed 30224 women aged 15–49 and 5021 men aged 15–59. For the purpose of this chapter, only respondents with known values on marital status at the time of the survey as well as on their education and residence are used in the analyses. This resulted in 10238 usable records for women and all 5021 records for men. Detailed tabulations for the entire survey may be found in the EPHS2010 final report, National-Statistics-Office-Eritrea and Fafo-AIS (2013). Summary statistics for the subset of data used in the present chapter are shown in Table 1.

By the survey time, 7421 of the 10238 women (72%) and 2569 of the 5021 men (51%) have responded they were ever married (this includes those who might have been separated or widowed after). The rest, 2817 women and 2452 men (28% and 49%, respectively), have responded that they were still single at the time of interview. The distribution of the women across educational levels shows that 4186 (41%) had no education at all, 2055 (20%) had primary-level education, 1827 (18%) had middle-level education, 1894 (18%) had secondary-level education, while the rest 276 (3%) had post-secondary education. The corresponding figures for men are 1051 (21%), 803 (16%), 1209 (24%), 1516 (30%), and 442 (9%), respectively. Further, 1819 (18%) of the women respondents were from the capital (Asmara), 2504 (24%) were from other towns, while the majority 5915 (58%) were from rural areas. The corresponding figures for men are 931 (19%), 1257 (25%), and 2833 (56%), respectively.

The columns of percentage married in Table 1 reveal clear differentials across both educational levels and residence for both women and men. For instance, while women with no education constitute 41% of the entire sample, they constitute 51% of the marriages (3799 of 7421). Women with post-secondary education, on the

other hand, constitute only 1% of the marriages (96 of 7421). The pattern is similar but less dramatic for men—those with no education constitute 35% of the marriages, while those with post-secondary education constitute only 8% of the marriages. Differentials across residence show that women from rural areas constitute 58% of the sub-sample but 64% of the marriages. Women from the capital, on the other hand, constitute 19% of the sub-sample but only 13% of the marriages. The contribution of men from Asmara to the sub-sample is 18%, while their contribution to the total marriage is 15%. Men from rural areas constitute 56% of the sub-sample but 65% of the marriages.

Plots of probabilities of family initiation by age and across educational levels and residential areas are shown in Fig. 1. The upper panel are educational (left) and residential (right) plots for women. Those in the middle panel are corresponding plots for men. The plot in the lower panel shows gender differences in the probability of family initiation at each age.

The plots indicate that there are differentials in the probabilities of family initiation across education and residence and that the educational differences are more pronounced among women (upper panel) than among men (middle panel). Further, the last plot (lower panel) shows that women initiate family at much faster (at younger ages) than men.

These summary tables and figures provide some overview of the data at hand, but in order to better understand the differentials and make well-grounded inferences, we need to support our initial observation with appropriate analyses of the data and formal statistical tests. Ghilagaber (2018) has analyzed the data sets using frequentist statistical methods ranging from elementary measures of association between marriage and the covariates to more complex and advanced methods that utilize the data more efficiently. The data sets are also analyzed elsewhere in this book in a Bayesian accelerated failure-time framework with the extended generalized gamma model and its 13 special cases, see Liang and Ghilagaber (2022).

For continuity and comparison with later sections, we present in Table 2 results from fitting the standard Cox proportional hazards model in Eq. (1) separately for women, men, and the combined sample.

The separate results for women and men are in accordance with the findings in Ghilagaber (2018), but the combined sample was not analyzed before.

In Fig. 1, we observed that the educational and residential differentials in probabilities of family initiation are not uniform across the ages for both men and women. The last plot in Fig. 1 also indicates that even the gender difference in the probabilities is not uniform across ages.

The above observations justify re-analysis of the data using models that account for non-constant differentials. In the next section, we present our proposed dynamic model in a Bayesian framework. The presentation of the model will be relatively brief and will take up topics that are relevant in order to understand the application in Sect. 4. A more detailed description of the model and its associated efficient particle filter used to sample from the posterior distribution can be found in Munezero (2022).

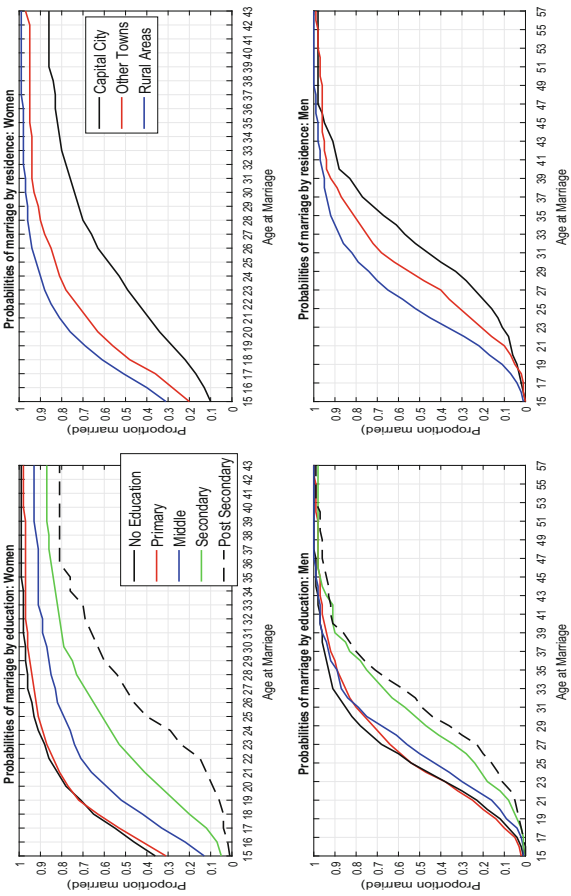


Fig. 1 Probabilities of marriage by age and across education and residence for men and women

Table 2 Estimated relative intensities of marriage from standard Cox PH models

		Women			Men			Combined sample		
Covariate	Levels	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
Education	No Educ	4.42	3.59	5.45	1.54	1.30	1.81	3.97	3.52	4.48
	Primary	4.31	3.49	5.31	1.85	1.57	2.20	3.57	3.16	4.03
	Middle	2.78	2.25	3.44	1.61	1.36	1.91	2.26	2.00	2.56
	Sec	1.82	1.47	2.25	1.13	0.96	1.33	1.44	1.27	1.63
	PostSec	1	–	–	1	–	–	1	–	–
Residence	Capital	1	–	–	1	–	–	1	–	–
	Other Towns	1.42	1.31	1.55	1.33	1.16	1.53	1.26	1.17	1.35
	Rural Areas	1.57	1.44	1.70	1.89	1.65	2.17	1.28	1.20	1.37

3 Dynamic Bayesian Modeling of Survival Data

Following notation in Munezero (2022), we denote the random survival time by $\tilde{T}$. It represents the time until the event of interest (marriage) occurs to an individual or the study period ends (the individual is censored)—whichever comes first. Denoting the censoring variable by C, the observed time is represented by the random variable $T = \min(\tilde{T}, C)$.

The *hazard function* describes the instantaneous rate at which the event (marriage) occurs and is linked to $\mathbf{x}$ as in Eq. (2).

The corresponding *survival function* is then given by

$$S(t|\mathbf{x}) = \exp\left(-\int_0^t \lambda(s|\mathbf{x})ds\right) \tag{3}$$

and is defined as the probability that an individual with profile $\mathbf{x}$ has not experienced the event by time t.

Thus, given observed exposure times for n individuals, t_1,..., t_n, a censoring indicator d_i ($d_i = 0$ for censored, $d_i = 1$ for events), and the covariates $\mathbf{x}_i$ (for $i = 1, \cdots, n$), the resulting likelihood function is given by

$$L(t_1, \ldots, t_n|\boldsymbol{\beta}(t)) = \prod_{i=1}^{n} \lambda(t_i|\mathbf{x}_i)^{d_i} S(t_i|\mathbf{x}_i), \tag{4}$$

where the dependence of $\lambda(t|\mathbf{x})$ on $\beta(t)$ is as given in Eq. (2).

Below, we redefine the likelihood in the framework of piece-wise constant hazard model in Gamerman (1991). The observation time is partitioned into smaller intervals, and the likelihood is obtained as the product of the likelihoods in each interval.

3.1 The Likelihood Under Piece-Wise Exponential Distribution

In piece-wise exponential models, the exposure time is partitioned into consecutive disjoint intervals, $I_j = [\tau_{j-1}, \tau_j)$ (where $j = 1, \cdots, J$ and $\tau_0 = 0 < \tau_1 <, \cdots, < \tau_J$). The baseline hazard function is assumed to be constant within each interval I_j, i.e., $\lambda_0(t) = \lambda_{0j}$, for $t \in I_j$ and $\lambda_{0j} > 0$. Further, it is assumed that the vector of coefficients $\boldsymbol{\beta}(t)$ is piece-wise constant, i.e., $\boldsymbol{\beta}(t) = \boldsymbol{\beta}_j$ if $t \in I_j$. Thus, the hazard function is represented by several constant parameters $\lambda_1, \ldots, \lambda_J$, where each λ_j is connected to the covariate information of an individual i through the log link

$$\ln \lambda_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j, \tag{5}$$

which allows flexibility to capture different shapes of the hazard function across time.

In Eq. (5), $\mathbf{x}_i$ is the original covariate vector augmented with a column of 1, and $\boldsymbol{\beta}_j$ represents the vector of regression coefficients, where the intercept $\beta_{0j} = \ln(\lambda_{0j})$ is the log of the baseline hazard.

Partitioning time into discrete intervals breaks the survival time t_i into several exposure times $t_{ij} = \max(0, \min(t_i - \tau_{j-1}, \tau_j - \tau_{j-1}))$, which define the length of time individual i is exposed to the event of interest within the interval I_j. The exposure time is equal to the length of I_j (if individual i survived through this interval), or it is equal to $t_i - \tau_{j-1}$ (if individual i experienced the event within interval I_j); otherwise, it is equal to zero. Similarly, the event indicator expands into a vector of binary variables $d_{ij} = 1$ if individual i experiences the event in interval I_j, and $d_{ij} = 0$ if individual i survives through interval I_j.

Assuming covariates enter the model as in Eq. (5), the survival function for individual i becomes

$$S(t_i|\mathbf{x}_i) = \exp\left(-\left[\sum_{j=1}^{h-1} \lambda_{ij}\left(\tau_j - \tau_{j-1}\right)\right] - \lambda_{ih}\left(t_i - \tau_{h-1}\right)\right), \text{ if} \tau_{h-1} \le t_i < \tau_h,\ h \le J, \tag{6}$$

and the likelihood in Eq. (4) can now be factorized across the intervals:

$$L\left(\mathbf{t}_{1:J}|\boldsymbol{\beta}_{1:J}\right) = \prod_{j=1}^{J}\left[\prod_{i=1}^{n_j} \lambda_{ij}^{d_{ij}} \exp\left(-\lambda_{ij} t_{ij}\right)\right], \tag{7}$$

where $\mathbf{t}_j$ is the vector of exposures for interval I_j, $\mathbf{t}_{1:J} = (\mathbf{t}_1, \ldots, \mathbf{t}_J)$, $\boldsymbol{\beta}_{1:J} = (\beta_1, \ldots, \beta_J)$, and n_j is the number of individuals who experienced the event in interval I_j.

3.2 Prior Specification

We use one of the simplest and most common smoothing priors on the regression coefficients, $\boldsymbol{\beta}_j$. This simple prior is the random walk:

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_{j-1} + \epsilon_j, \quad \epsilon_j \sim N\left(0, \mathbf{U}_j\right). \tag{8}$$

This process is a special case of the more general first-order random walk process for parameter evolution suggested by Gamerman (1991) and has been used by Hemming and Shaw (2002) and Wagner (2011), among others.

It is clear from Eq. (8) that if $\mathbf{U}_j$ is a zero matrix, then there is no change in the regression coefficients over time, and, thus, the dynamic model reduces to the standard proportional hazards model. Else, $\boldsymbol{\beta}_j$ varies over time, and larger values for the entries in $\mathbf{U}_j$ induce high variations in $\boldsymbol{\beta}_j$.

In many applications, $\mathbf{U}_j$ is assumed constant, and it can be a diagonal matrix or a full matrix. For a diagonal $\mathbf{U}_j$, appropriate priors for diagonal elements are the log-normal priors according to Hemming and Shaw (2002) or the inverse gamma priors according to Sargent (1997) and Wagner (2011). For a full matrix, an inverse Wishart prior is adequate according to Gamerman (1998).

Alternatively, West et al. (1985) argue that it is possible to avoid computing $\mathbf{U}_j$ by using a discounting procedure to approximate $\mathbf{U}_j$ adaptively. This procedure uses a discount parameter $0 < \phi < 1$ that controls the amount of information transferred through intervals. Assuming that Σ_{j-1} is the posterior variance of parameters in the previous interval $j-1$, then $\mathbf{U}_j = (\phi^{-1} - 1)\Sigma_{j-1}$. Therefore, a discount factor close to 1 penalizes high fluctuations making the parameters evolve in a static way. Else, parameters are allowed to move freely and adapt to local changes over time. For more details, see Munezero (2022) and West et al. (1985).

3.3 Sampling and Inference from the Posterior Distribution

Combining the likelihood in Eq. (7) and the prior in Eq. (8) yields the following posterior distribution:

$$p\left(\boldsymbol{\beta}_{1:J}|\mathbf{t}_{1:J}\right) \propto p\left(\mathbf{t}_1|\boldsymbol{\beta}_1\right) p\left(\boldsymbol{\beta}_1\right) \prod_{j=2}^{J} L_j\left(\mathbf{t}_j|\boldsymbol{\beta}_j\right) p\left(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{j-1}\right), \tag{9}$$

where $p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{j-1})$ is defined by the expression in Eq. (8).

Markov chain Monte Carlo (MCMC) methods have been used to sample from the posterior in Eq. (9) in, for instance (Frühwirth-Schnatter, 1994; Hemming and Shaw, 2002; Wagner, 2011).

But, since the likelihood in Eq. (7) and the prior process in Eq. (8) define a state space model with nonlinear and non-Gaussian observation model, it is appropriate to use sequential Monte Carlo (SMC) methods for inference as argued by Gordon et al. (1993). Further, according to Carpenter et al. (1999); Doucet et al. (2000, 2001) SMC methods are specifically designed for filtering problems in state space models, where the main objective is to sample from $p(\boldsymbol{\beta}_{1:j}|\mathbf{t}_{1:j})$, $j = 1, \ldots, J$, sequentially through lower-dimensional filtering distributions, $p(\boldsymbol{\beta}_j|\mathbf{t}_{1:j})$. More detailed description of SMC can be found in Munezero (2022).

4 Application: Dynamic Bayesian Modeling of Time to Family Initiation

We now apply the dynamic model described in Sect. 3 on the data described in Sect. 2. Our proposed dynamic model requires partitioning the time variable (age) into J intervals. As described in Gamerman (1991), the common practice is to set the interval limits at each event time though such procedure can, sometimes, lead to too many intervals. Our time variable in the present illustration, age, is relatively simple, and thus we have partitioned it by 1 year intervals from age 15 until the highest observed age at marriage. This resulted in 27 intervals for women (with observed marriages at all ages between 15 and 43). The corresponding number of intervals for men was 40 (ranging from ages 15 to 56).

Results from fitting our dynamic model to our data sets are presented in Fig. 2. The upper-left panel of Fig. 2 shows relative intensities of entry into first marriage across educational levels among women. Women with the highest educational level (post-secondary education) were used as baseline (reference) level. We see clearly in figure that educational differences in entry into first marriage are not constant across woman's age. The differences are clearly distinguishable in the young ages (until about age 22) where women with no or lower education are more likely to enter into marriage than those with higher education but with varying degrees. After age 22, the differences diminish (and even go in the opposite direction), and the negligible differences are constant over ages.

The upper-right panel of Fig. 2 shows relative intensities of entry into first marriage across educational levels among men. Men with the highest educational level (post-secondary education) were used as baseline (reference) level. The figure again shows clearly that educational differences in entry into first marriage are not constant across men's age. Further, the patterns of variation across ages are different from those of women. In the figure for men, we see that the variations in relative intensities continue until about age 30, then stabilize between ages 30 and 40, and again begin to show variations after age 40. At young ages (until about age 25), education is negatively related to entry into first marriage. After that, the educational differences begin to diminish or, in fact, change direction. For instance, after age 30,

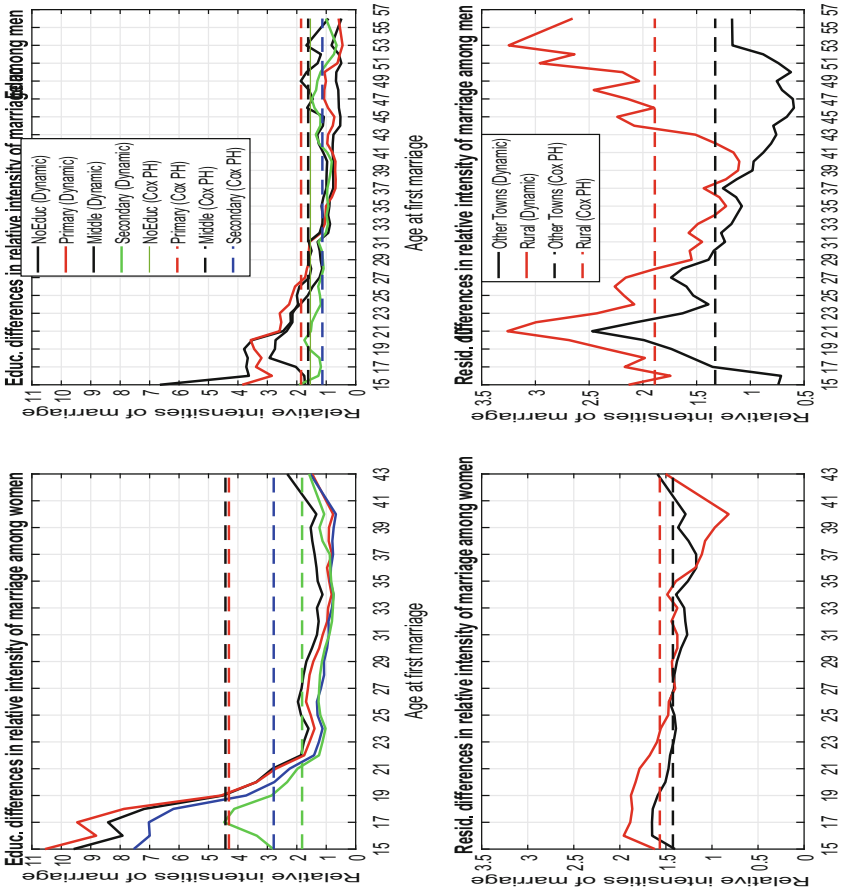


Fig. 2 Relative intensities of marriage across education and residence for women and men

it is the curve for men with middle-level education that lies at the top of all curves although it is not far from that of the baseline.

Residential differences in intensities of first marriage are shown in the lower panel of Fig. 2 for both women (left) and men (right). Women and men from the capital city (Asmara) are used as baseline levels. For women, residential differentials are well pronounced at young ages (until about age 27) and disappear after that. For men, on the other hand, the residential differences are pronounced at early ages (until about 30) and at old ages (after age 40) but seem to be stable and insignificant between age 30 and 40.

A result worth noting in the lower panel of Fig. 2 is that men from rural areas have higher intensities of entry into first marriage than those from the capital (Asmara) at all ages. The relative intensities of men from other towns oscillate around 1 indicating smaller differences from those of the capital. For women, on the other hand, those from rural areas have higher intensities than those from the capital at younger ages, but the differences get smaller over time and disappear at older ages. Women from other towns have relative intensities that stay steadily above those from the capital at all ages.

To make the gender differences in the relative intensities more clear, we present, in Fig. 3, pairs of relative intensities of the six covariates studied (four educational levels and two residential areas). We thus plot the relative intensities by education (upper and middle panels) and residence (lower panel).

In Table 3, we present the posterior means of relative intensities (averaged over the 27 intervals for women and 40 intervals for men) and their corresponding 95% credible intervals.

We note in Table 3 that the only significant relative intensities (whose 95% credible interval does not include 1) are women with no education (relative to women with post-secondary education) and those from other towns (relative to women from the capital city) as well as men from rural areas (relative to men from the capital city).

These results are in stark contrast with those in Table 2 where the relative intensities of all educational and residential groups were significantly different from their respective baselines. But, the results in Table 3 are in accordance with the behavior of the corresponding curves in Fig. 2. The figures show significant differences in younger ages but no difference (or difference in the opposite direction) over the larger portion of the age interval. Thus, when these differences are averaged, it should not come as a surprise if they do not show any significant difference between some of the study groups.

We sum up our results in Table 4 where we reproduce the relative intensities of entry to first marriage from the dynamic model (Table 3) and the static Cox proportional hazards model (Table 2). We note that the relative intensities in the dynamic model are much lower than their corresponding entries from the Cox proportional hazards model. In the third columns, we provide percentage differences between the estimates from the models.

We see that estimates of relative intensities among women's educational groups are much higher in the Cox PH model than in the dynamic model. The same is true

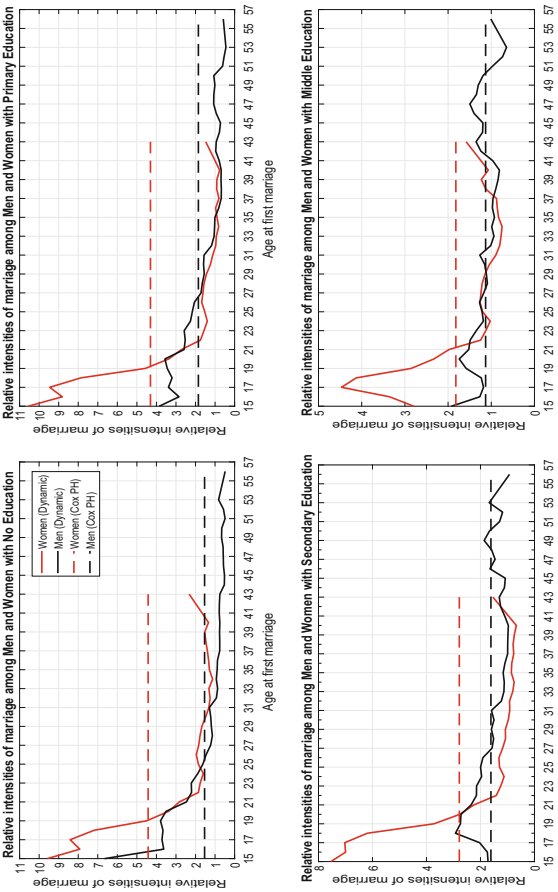


Fig. 3 Gender differences in relative intensities of 4 educational levels and 2 residential areas

Table 3 Estimated posterior means of relative intensities of marriage and their 95 % credible intervals from dynamic Bayesian models

		Women			Men		
Covariate	Levels	Post. mean	2.5%	97.5%	Post. mean	2.5%	97.5 %
Education	No Educ	2.17	1.15	9.36	1.13	0.49	5.22
	Primary	1.79	0.77	10.36	1.29	0.49	3.71
	Middle	1.48	0.69	7.45	1.50	0.95	2.83
	Sec	1.38	0.75	4.41	1.15	0.69	1.83
	PostSec	1	–	–	1	–	–
Residence	Capital	1	–	–	1	–	–
	Other Towns	1.40	1.17	1.65	1.11	0.60	2.27
	Rural Areas	1.45	0.85	1.95	1.89	1.11	3.25

Table 4 Relative intensities of marriage from dynamic models (posterior means) and Cox PH models: replicated from Tables 2 and 3, respectively, for comparison

		Women			Men		
Covariate	Levels	Dynamic	Cox PH	Diff. (%)	Dynamic	Cox PH	Diff. (%)
Education	No Educ	2.17	4.42	103.69	1.13	1.54	36.28
	Primary	1.79	4.31	140.78	1.29	1.85	43.41
	Middle	1.48	2.78	87.84	1.50	1.61	7.33
	Sec	1.38	1.82	31.88	1.15	1.13	1.77
	PostSec	1	1	–	1	1	–
Residence	Capital	1	1	–	1	1	–
	Other Towns	1.40	1.42	1.43	1.11	1.33	19.82
	Rural Areas	1.45	1.57	8.28	1.89	1.89	0.00

for men, but the percentage differences between the models are much lower than those in the women data. In fact, the estimates of relative intensities for men with secondary-level education are practically the same (1.15 and 1.13, respectively).

For residence, we see that the two models yield practically the same estimates of relative intensities for women from other towns and identical estimates for men from rural areas.

5 Summary and Concluding Remarks

In this chapter, we presented a dynamic Bayesian survival model that relaxes the assumption of proportional hazards in conventional models and allows effects of covariates to vary over time.

The model was described in the framework of piece-wise exponential distribution (piece-wise constant hazards) where the observation time was partitioned into small intervals. Further, a simple random walk process was assumed as prior for

the coefficients, and information was transferred between intervals via a discount parameter. Sequential Monte Carlo method, also known as particle filter, was used to sample from the posterior distribution.

We illustrated the proposed model by fitting it to data on family initiation (entry to first marriage) among men and women from Eritrea based on its 2010 Population and Health Survey. Our empirical results show that educational and residential differentials in family initiation are not uniform over time (in our case age of respondents). Thus, we argue that the dynamic survival model that allows covariate effects to vary over time captures patterns of family initiation more correctly.

We also found that educational and residential differentials are significant at younger ages but diminish after a short period and are insignificant for a longer part of the observation period. This, in turn, implied that the average values of the relative intensities over the entire period were insignificant for most of the covariates. In fact, it was only women with no education and those from other towns, as well as men from rural areas that differed significantly from their respective baselines. This was in stark contrast with the results from the conventional proportional hazard models where all covariates were significant.

Our results also indicated that the patterns of variation in the relative intensities are different between men and women. For instance, educational differences are more pronounced among women than among men, while residential differences are much smaller than those of education for both men and women.

We are aware that there can be interaction between the two covariates used in the analyses (education and residence). Individuals with higher education are more likely to live in the capital city or other towns than in rural areas. Thus, the estimated effects may not reflect pure effects of the respective covariates.

Further, the educational level used in the analyses is what was achieved by the survey date (2010), while the event of interest (marriage) might have taken place earlier. This is a common problem in retrospective data. If a high proportion of respondents have completed the reported educational level after they have married, the educational differentials in marriage intensities may be distorted due to misclassification of individuals over educational levels. Ghilagaber and Koskinen (2009), Ghilagaber and Larsson (2019), and Munezero and Ghilagaber (2022) study potential biases due to use of anticipatory covariates and how to account for that.

Last, the event of interest in the illustration, entry into first marriage, is not a certain event (like, for instance, death). There may be individuals who, for various reasons, may never marry. Such individuals are known long-term survivors in the literature. Thus, among the individuals who were not married at the time of interview, some of them are "genuinely censored" and may marry sometime after the survey, while others may never marry.

One possible direction for future work may, therefore, be to use mixture models to re-analyze the data. Among the censored individuals, a logistic regression model may be used for the probability of being long-term survivor (never marry). This model may then be estimated jointly with the model for the intensity of marriage among all individuals.

Acknowledgments The data sets analyzed in this chapter were provided to the corresponding author (GG) by the National Statistics Office, Eritrea. He is grateful to its director Mr. Ainom Berhane and its senior statisticians Mr. Hagos Ahmed and Mr. Samuel Tesfamariam. The views expressed in the chapter are solely of the authors and do not express the views or opinions of the data source or its employees.

Appendix: R Program Codes for Dynamic Bayesian Survival Modeling Used in the Chapter

```
#-----------------------------------------------------------------
# Load required file
# ------------------------------------------

# ------------------------------
# extract data and cleaning data
# ----------------
# function to create dummy variables
dummies <- function(data, variables,ref){
  cols <- colnames(data)
  new_var_name <- cols[which(!(cols %in% variables))]
  new_data <- data[new_var_name]
  for (i in seq_len(length(variables))){
    fctr <- sort((unlist(unique(data[variables[i]]))))

    levels <- fctr[-which(fctr==ref[i])]
    for(k in levels){
     new_var_name <- c(new_var_name,paste(variables[i],k,sep=""))
     new_data <-cbind(new_data, ifelse(data[variables[i]]==k,1,0))
    }
  }
  colnames(new_data)<-new_var_name
  return(new_data)
}
colnames(data)<-c("age","marriage_ind","birth_cohort","residence",
"education")View(data)
data$education <- data$education+1
train_data <- dummies(data[,-3], variables=c("residence",
"education"), c(1,5))
# cohort discarded from analysis
var_names <- c("intercept", colnames(train_data)[-c(1,2)])
head(train_data)
#-----------------------------------------------------------------
# setting the initial values
#-----------------------------------------------------------------
m_init<-rep(0,dim(train_data)[2]-1) # initial mean of the
regression parameters c_init<-diag(10,dim(train_data)[2]-1)
# initial convariance matrix for the regression parameters
event_time<-unique(sort(train_data[which(train_data[,2]==1),1]))
# event times intervals<-c(event_time,max(train_data[,1]))
```

```
# fitting the model
results<-SmoothFilter(train_data,intervals,M=2500,
                      c_init,R=2,alpha=.45,T_ind=1,C_ind=2)
particles<-results$Particles
Pf.sumary<-trace.summary(particles)
PfMean.trace<-Pf.sumary$Mean
pfHPD<-Pf.sumary$HPD

# exprort the posterior sample
saveto <- file.path(getwd(),paste("women_2010_","posterior_",Sys.
Date()))apply(as.matrix(1:length(var_names)),1,
      function(f) write.csv2(particles[[f]], paste(saveto,
      var_names[f],".csv")))
# plot the parameter evolution

library(shape)
color<-shadepalette(n = 10, "white", "black")
col.names<-c("Log␣baseline",colnames(train_data)[-c(1,2)])
windows(18,10,0.7)
plot.new()
par(mfrow=c(2,(length(particles)+1)/2),font=1,family="sans")
m=1
l=length(intervals)
xmax<-ceiling(max(train_data[,1]))
xaxis<-intervals
for(i in 1:length(particles)){

  plot(xaxis,rep(range(pfHPD[c(m,m+1),]),length.out=length(xaxis)),
       type="n",ylab="␣", xlab="␣",main="␣",axes=FALSE) # axes
       suppressed

  polygon(c(rev(rep(xaxis,each=2)[-c(1,2*l)]), rep(xaxis,each=2)
  [-c(1,2*l)]),c(rev(rep(pfHPD[m,],each=2)),rep(pfHPD[m+1,],
  each=2)), col =color[9],border = color[9])

  lines(xaxis,c(PfMean.trace[,i],PfMean.trace[l-1,i]),type="s",
  lty=1,lwd=2,col="black") abline(h=0,lty=2,lwd=2)
  ygrid<-round(range(pfHPD[c(m,m+1),]),2)
  xgrid<-intervals
  ylablist<-round(seq(ygrid[1],ygrid[2],by = round((ygrid[2]
  -ygrid[1])/10,2)),2)
  axis(1, at=xgrid, labels = xgrid,las=1,tck=-0.02,cex.axis=1,
  font=2)
  axis(2,at=ylablist,labels=ylablist,las=2,tck=-0.03,
  cex.axis=1,font=2)

  # add text
  mtext("Time␣in␣years",1,cex=1,line=3,font=2)
  mtext(col.names[i],cex=0.8,line=.7,font=2)
  box(lty=1) # surounding box
  m<-m+2
}
```

References

Carpenter, J., Clifford, P., & Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation, 146*(1), 2–7.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, 34*, 187–220.

Doucet, A., De Freitas, N., & Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In Sequential Monte Carlo methods in practice (pp. 3–14). Berlin: Springer.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing, 10*(3), 197–208.

Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika, 81*(2), 317–330.

Fahrmeir, L., & Kneib, T. (2011). *Bayesian smoothing and regression for longitudinal, spatial and event history data*. New York: Oxford University Press.

Fahrmeir, L., & Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association, 91*(436), 1584–1594.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis, 15*(2), 183–202.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 40*(1), 63–79.

Gamerman, D. (1998). Markov Chain Monte Carlo for dynamic generalised linear models. *Biometrika, 85*(1), 215–227.

Ghilagaber, G. (2018). Statistical analysis of educational and urban-rural differences in marriage-timing: Evidence from the 2010 Eritrean population and health survey for women and men. In *Proceedings of the International Conference on Eritrean Studies, 20–22 July 2016* (pp. 1097–1124).

Ghilagaber, G., & Koskinen, J. H. (2009). Bayesian adjustment of anticipatory covariates in analyzing retrospective data. *Mathematical Population Studies, 16*(2), 105–130.

Ghilagaber, G., & Larsson, R. (2019). Maximum likelihood adjustment of anticipatory covariates in the analysis of retrospective data. Research report 2019-1. Stockholm: Department of Statistics, Stockholm University.

Godha, D., Gage, A., Hotchkiss, D., & Cappa, C. (2016). Predicting maternal health care use by age at marriage in multiple countries. *Journal of Adolescent Health, 58*(5), 504–511.

Gordon, N., Salmond, D., & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)* (Vol. 140, pp. 107–113). New York: IET.

Hemming, K., & Shaw, J. (2002). A parametric dynamic survival model applied to breast cancer survival times. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 51*(4), 421–435.

Hemming, K., & Shaw, J. (2005). A class of parametric dynamic survival models. *Lifetime Data Analysis, 11*(1), 81–98.

Hennerfeind, A., Brezger, A., & Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association, 101*(475), 1065–1075.

Liang, Y., & Ghilagaber, G. (2022). Bayesian survival analysis with the extended generalized gamma model: application to demographic and health survey data. In *Modern Biostatistical Methods for Evidence-Based Global Health Research*. Springer Series on Emerging Topics in Statistics and Biostatistics.

Munezero, P. (2022). Efficient particle smoothing for Bayesian inference in dynamic survival models. *Computational Statistics, 37*(2), 975–994.

Munezero, P., & Ghilagaber, G. (2022), Dynamic Bayesian adjustment of anticipatory covariates in retrospective data: application to the effect of education on divorce risk. *Journal of Applied Statistics, 49*(6), 1382–1401.

National-Statistics-Office-Eritrea and Fafo-AIS (2013). *Eritrea Population and Health Survey 2010*. National Statistics Office and Fafo Institute for Applied International Studies.

National-Statistics-Office-Eritrea and Macro-International-Inc. (1997). *Eritrea Demographic and Health Survey 1995*. New York: National Statistics Office and Macro International.

National-Statistics-Office-Eritrea and Macro-International-Inc. (2003). *Eritrea Demographic and Health Survey 2002*. National Statistics and Evaluation Office and ORC Macro.

Raj, A. (2010). When the mother is a child: the impact of child marriage on the health and human rights of girls. *Archives of Disease in Childhood, 95*(11), 931–935.

Sargent, D. (1997). A flexible approach to time-varying coefficients in the Cox regression setting. *Lifetime Data Analysis, 3*(1), 13–25.

van Houwelingen, H., & Putter, H. (2011). *Dynamic prediction in clinical survival analysis*. New York: CRC Press.

Wagner, H. (2011). Bayesian estimation and stochastic model specification search for dynamic survival models. *Statistics and Computing, 21*(2), 231–246.

West, M., Harrison, P., & Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association, 80*(389), 73–83.

Bayesian Spatial Modeling of HIV Using Conditional Autoregressive Model

Ropo Ebenezer Ogunsakin and Ding-Geng (Din) Chen

Abstract *Background:* In the spatial analysis, the conventional method for disease modeling and mapping is based on a log-linear relationship between relative risk and local variation, while the covariates are ignored. On the other hand, the general assumption in spatial modeling is the stationarity of the mean, which implies the associations between the relative risk and some set of covariates, which is constant over regions. In reality, the comparative risk modeling usually infringes on this stationarity assumption because of spatial dependencies. Thus, non-stationarity of the mean can be employed using the Spatially Varying Coefficients (SVCs) model. *Method:* In this study, we propose a generalized linear model (GLM) with Bayesian inference to build the SVC model and compared it with the stationary model. The SVC model is used to relax the stationarity assumption in which nonlinear effects of age are captured through the random walk of order two and by allowing the covariates to vary spatially using a conditional autoregressive model. This study aimed to profile people living with HIV in Nigeria. In this chapter, identical spatial regression models are fitted for Bayesian approach, using General Household Survey (GHS) data for the year 2015. *Result and Conclusion:* The finding of this study highlights a nonlinear relationship between the incidence of HIV and age. Among others, this study highlights areas where women are at higher risk of HIV infection across the six regions of Nigeria. The modeling of the socio-demographic predictors of HIV infection and spatial maps provided in this study could aid in developing a framework to alleviate HIV and identify its hotspots for urgent intervention in the endemic regions.

R. E. Ogunsakin (✉) · D.-G. (Din) Chen
University of Pretoria, Pretoria, South Africa
e-mail: din.chen@up.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_13

1 Introduction

Human Immunodeficiency Virus (HIV) continues to be a major threat in the global community. It is estimated that 36.9 million people were living with the virus at the end of 2017 (Armstrong et al., 2018), of which 25% of these same people do not know that they have the virus. The vast majority of people living with HIV and acquired immune deficiency virus (AIDS) are found in low- and middle-income countries, with an estimated 66% of them living in Sub-Saharan Africa (Bekker et al., 2018). Out of these, it is estimated that 19.6 million are resided in either East or Southern Africa. Further statistics put it that around two-thirds of new HIV infections in West and Central Africa in 2017 occurred in Nigeria (Lamont et al., 2012). Moreover, this region also estimated to have the highest prevalence and HIV infection incidence in the world (Beyrer et al., 2016).

Several mechanisms have been developed by Nigeria government and non-governmental organizations to reduce new infections as well as improve the living standard of people infected with HIV. Some of the measures devised are the roll out of antiretroviral therapy (ART) (AIDSinfo, 2016). Hence, the control of HIV prevalence is necessary especially for assessing the trend as well as comparison between regions.

The first HIV infection in Nigeria was reported in 1986, and since then, the prevalence has risen from less than 0.1% in 1987 to 5.8% in 2002 with the estimated 3.6 million Nigerians (EO et al., 2005). The national HIV prevalence rate in 2012 was estimated to be 3.4% among adults aged 15–49 years (Awofala & Ogundele, 2018) out of 3.1 million people living with HIV, with a regional variation. The prevalence of HIV varies considerably between the different geopolitical zones of Nigeria. The southern Nigeria (South Zone) recorded the highest HIV prevalence at 5.5%, while the lowest prevalence rate is in the southeast (South-East Zone) with a prevalence of 1.8% (Bekker et al., 2018). The total prevalence rate of HIV is higher in rural areas (4%) than in urban (3%) (Baral et al., 2012; Aniekwu, 2002). The differences in the prevalence across the geopolitical zones can be attributed to barriers to the HIV response, which can be explained by cultural barriers, structural barriers, and economic barriers (Group, 2003; Salaam-Blyther & Kendall, 2012; WHO, 2015). Therefore, the clear understanding on the spatial distribution which can bring about an accurate disease modeling and mapping to prevent the spread of the disease across the different regions is needed.

In previous studies, (Djukpen, 2012) reported that there is a significant spatial clustering of HIV/AIDS based on the mapped Global and Local Moran's I values, but the study did not examine the nonlinear relationship of age with HIV prevalence. In addition, the spatial distribution of HIV considered did not take into consideration the effect of some socioeconomic factors on HIV prevalence. Other studies outside Nigeria have found that HIV prevalence increases with age (Niragire et al., 2015; Ngesa et al., 2014), and in particular there is a nonlinear relationship between age and HIV prevalence, but this has not been considered in the context of Nigeria.

Considering a normal generalized linear regression model which has been utilized in spatial data analysis, most studies have assumed stationarity of the mean and covariance. By stationarity of the mean, it suggests a constant association between outcome of interest and a set of covariates over the region or zone. This assumption is unrealistic because of spatial dependencies and unknown factors that may impact the outcome. Thus, this assumption can only be realistic when the regression coefficients vary across space (Pearce, 1999). Therefore, the problem of non-stationarity can be accommodated by allowing the relationships measured to vary over space via geographically weighted regression (GWR) model or spatially varying coefficients parameter (SVCP) (Gelfand et al., 2003). Therefore, this study aims to perform a spatial analysis modeling to capture this nonlinearity of some covariates among people living with HIV and Bayesian inference are applied with Integrated Nested Laplace Approximation (INLA) to build SVC model. Specifically, this study employed a Bayesian spatially varying coefficients process (BSVCP) by assigning its coefficients with the conditional autoregressive (CAR) model while relaxing the stationarity as well as the linearity assumption using Demographic Health Survey data. Most countries carry out Demographic and Health Surveys (DHSs) in order to understand people's comprehension of certain health issues and also determine prevalence and awareness about several diseases. The survey data are usually collected with an aim of being representative of the whole population.

The rest of the chapter is organized as follows. In Sect. 1, we provide details about the data we collected in Nigeria, and in Sect. 2, we introduce the Bayesian spatial varying model with its statistical implementation and computation on priors and model selection criteria. Furthermore, we summarize the results of our data analyses in Sect. 3. Finally, Sect. 4 features discussions of our findings and concluding remarks.

2 Data and Models

2.1 Data Description

The data used for this study were obtained from Nigeria Demographic Health Survey (NDHS) conducted in 2015. It is the most recent population-based and nationally representative sample survey that tested for HIV infection. The resources for the conduct of the survey were provided by the United States Agency for International Development (USAID), United Nations Population Fund (UNFPA), United Kingdom Department for International Development (DFID), and Government of Nigeria through National Population Commission (NPC). The primary objective of the 2015 NDHS was to obtain relevant information on the prevalence of HIV among adults and to provide up to date information on attitudes regarding HIV.

The sample for the survey was nationally representative which covered the entire population. The sample design is allowed for specific indicators to be calculated

for the six (6) different zones. The sample was selected using a stratified three stage cluster design. It comprises 904 clusters, of which 372 urban and 532 rural. A fixed sample take of 45 households were selected per cluster where both men and women aged 15–49 years were interviewed. The survey used two questionnaires: household questionnaire was used to collect the characteristics of each person listed and the individual questionnaire to collect information based on demographic characteristics as well as knowledge of HIV on men and women aged 15–49 years. The content of these questionnaires was obtained from MEASURE DHS programmes. For more information on NDHS 2015, the readers should see full report (Macro et al., 2014). Information from 3975 women aged 15–49 years who were tested for HIV infection was used in the analysis. The data can be accessed from the website http://www.dhsprogram.com/data/dataset_admin/login_main.cfm.

2.2 Theoretical Model

Generalized linear models (GLMs) and extensions give a uniform structure for investigating the relation between a response variate y_i and a vector $X_i = (x_{i1}, \ldots, x_{it})$ of covariates observed for $i = 1, \ldots,$ n individuals. Let y_{it} denote the HIV status for individual i in a state t. The response variable HIV status has a binary outcome coded as HIV positive and negative. Suppose $y_{it} = 1$ when individual i in state t is HIV positive and zero otherwise. The present study assumes the outcome variable (denoted by y_{it}) as a univariate Bernoulli given as $y_{it}|p_{it} \sim Ber(p_{it})$. If the set of covariates contains q categorical covariates G and r continuous covariates X, this is represented as $X_{it} = (x_{it1}, \ldots, x_{itr})', \quad G_{it} = (g_{it1}, \ldots, g_{itq})'$. Typically, the unknown mean response given as $\mathrm{E}(y_{it}) = p_{it}$ is related to the independent variable in a link function given as

$$\psi(p_{it}) = X'_{it}\beta + G'_{it}\pi$$

where β denotes an r-dimensional vector of regression coefficients for the continuous independent variables, while π represents a q-dimensional vector of regression coefficients for the categorical independent variables. The probability of woman i from state t testing HIV positive is denoted by $p_{it} = \mathrm{E}(y_{it})$, and $\psi(.)$ is a logit link function with logit $\psi(p_{it}) = \log\left(\frac{p_{it}}{1-p_{it}}\right)$.

To relax the strict linear relationship, we consider the nonlinear effects of the continuous covariates and spatial autocorrelation in the dataset; the present study considered both second order random walk (RW2) and convolution model. Therefore, the continuous covariates are modeled non-parametrically. Meanwhile, in the case of our study, RW2 model was employed to relax the linear predictor given as

$$\psi(p_{it}) = \sum_{k=1}^{r} f_k(x_{itk}) + f_{spatial}(s_t) + G'_{ij}\pi \tag{1}$$

where function $f_k(.)$, k =1,...,r, represent nonlinear effects for the continuous covariates and the state effect denoted by $f_{spatial}(s_t)$ represents the spatial effect of each zone. It can therefore be categorized into structured and unstructured (random) effect, which is represented as follows (Lawson, 2013; Alexander, 2011):

$$f_{spatial}(s_t) = f_{structured}(s_t) + f_{unstructured}(s_t) \tag{2}$$

Hence, the full model is given as

$$\psi(p_{it}) = \sum_{k=1}^{r} f_k(x_{itk}) + f_{str}(s_t) + f_{unstr}(s_t) + G'_{ij}\pi \tag{3}$$

The inference in this study is based on full Bayesian estimation techniques, and hence the prior distribution of the model was specified to all the parameters. Many of the previous studies (Fahrmeir & Tutz, 2013; Eilers & Marx, 1996; Currie & Durban, 2002) have established the approach of estimating the smooth function $f_k(.)$, but in the case of our study, random walk model was employed in estimating the smooth function $f_k(.)$. The approaches for estimating the smooth function $f_k(.)$ have been discussed in the literature by many studies (Eilers & Marx, 1996), and some of the methods commonly used include penalized regression spline, Markov Random Fields (GMRFs). In the present study, we adopted the random walk model for estimating the smooth function $f_k(.)$.

2.2.1 Bayesian Spatially Varying Coefficient Parameter (BSVCP) Model

Several studies have been published in the literature on the assumption that the relationship between outcome variables and independent variable is constant via study region (Carlin et al., 2014), but this assumption may not be realistic for the spatial processes due to some factors such as altitudes; culture can contribute to this fact. The two commonly spatially varying models are geographically weighted regression (GWR) and Bayesian spatially varying parameter (BSVP). In the present study, the focus is on the BSVP, which is used to relax the stationary assumption. The Bayesian spatially varying parameter model is used here to make inference. For detailed discussion on Bayesian spatial models to map vital rates, the readers are enjoined to read (Bernardinelli & Montomoli, 1992; Lawson et al., 1999; Gilks et al., 1995) for better understanding. In contrast to the previous studies, our study allows for covariates parameters to vary spatially.

The Bayesian spatially varying parameter model is a hierarchical in nature in which the distribution of the data is expressed as conditional on unknown

parameters. Based on the proposition of (Gelfand et al., 2003), spatially varying parameter (SVCP) model is given as

$$y_{it}|p_{it} \sim Ber(p_{it})$$

$$\psi(p_{it}) = X'_{it}\beta + G'_{it}\pi$$

Thus, the prior distribution for the regression coefficient parameters is represented as follows:

$$[\pi|\mu_\pi, \sum_\pi] = N(1_{n\times 1} \circledast \mu_\pi, \sum_\pi)$$

The vector $\mu_\pi = (\mu_{\pi 0}, \ldots, \mu_{\pi p})'$ contains the means of the regression coefficients terms. In addition, the prior on the regression coefficients accounts for the possible spatial dependence through the covariance $\sum_\pi$. In this study, the Bayesian spatial varying parameter was employed to relax the stationarity assumption, while the varying coefficient is achieved by specifying the priors for $\pi's$ as an aerial unit model such as simultaneously and conditionally autoregressive models (abbreviated as SAR and CAR). The SAR model is computationally easier for use with likelihood methods. In contrast, the CAR model is computationally easier for Gibbs sampling used with Bayesian model fitting, and in this regard is often used to incorporate spatial correlation via a vector of spatially varying random effects $\phi = (\phi_1, \ldots, \phi_k)'$ of k components follows a multivariate Gaussian distribution having mean zero and B as the inverse of the dispersion matrix. Hence, the density for ϕ is given by

$$p(\phi) = (2\pi)^{-p/2}|B|^{-1/2} exp\left\{\frac{1}{2}\phi' B\phi\right\} \tag{4}$$

The conditional distribution of one component in terms of the elements of matrix B is expressed as

$$p(\phi|\phi_{-i}) = exp\left\{\frac{-a_{ii}}{2}(\phi_i - \sum_{k=i}\frac{-a_{ik}}{a_{ii}}\phi_k)^2\right\} \tag{5}$$

This implies that $\phi_i|\phi_{-i} \sim \mathrm{N}(\frac{-a_{ik}}{a_{ii}}\phi_k, \frac{1}{a_{ii}})$. Let $\mathrm{C} = (c_{ik}) = \frac{-a_{ik}}{a_{ii}}$ and $\mathrm{D} = \mathrm{diag}$ $(\tau_1^2, \ldots, \tau i^2)$ such that $c_{ik}\tau_k^2 = c_{ki}\tau_i^2$. Hence, the inverse of the matrix Ω is related to C and M as given by

$$B = (I - C)D^{-1} \tag{6}$$

The joint distribution ϕ is *MVN* $(0, H^{-1}(I - G))$ provided $(I - H^{-1})G^{-1}$ is symmetric and positive definite, and I is the identity matrix (Besag, 1974). As noted by (Cressie & Wikle, 2015; Sherman, 2011; Carlin et al., 2014), the logic here is that

C and D must be modeled properly in order to ensure symmetry in Ω, while matrix C indicates the relationship between the neighbors. Particularly, the CAR model is an attractive way to handle spatial statistical dependencies (see (Cressie & Wikle, 2015; Carlin et al., 2014) for more details). Typically, the prior for the structured and unstructured random effects followed the CAR model and independently and identically distributed normal distribution, respectively. Hence, the specification of the Bayesian SVCP model in Equations (4) and (5) can be completed with the specification of the prior distribution of the parameters. Separate models were fitted and their deviance information criterion (DIC) values were compared. The DIC is a generalization of the Akaike's information criterion (AIC). A small DIC value corresponds with a good predictive performance of the model as defined by (Spiegelhalter et al., 2002), and it measures the fit and the complexity of each model. The fit of each model is measured by the posterior expectation $\bar{D} = E_{\phi|y}(\phi)$ of the deviance. The complexity is given by the effective number of parameters pD that is defined by the difference of the expected posterior deviance $\widehat{D}$. In addition, the deviance computed at the posterior $\widehat{\phi} = E_{\phi|y}(\phi)$ of the parameter, written as pD= $\widehat{D}$ - D($\widehat{\phi}$). Thus, pD is a penalty that penalizes a better fit by greater complexity (Spiegelhalter et al., 2002). Hence, the DIC is defined as

$$DIC = \bar{D}(\theta) + pD$$

The DIC is also regarded as one of the best approaches for comparing multi-level models, and its usage is very common in comparison of spatial–temporal Bayesian models. However, we observe that the DIC has been criticized by many researchers (Waller et al., 1997; Aitkin, 2010) and can be problematic in models with many random effects and thus should be used with care. It has been reported that a difference in DIC of 3 between two models cannot be distinguished and a difference of 3–7 has considerably less support (Spiegelhalter et al., 2002). We fitted the model using the R-INLA package (Martins et al., 2013) together with the R software (R Core Team, Vienna, Austria) (Team, 2016). The models under consideration are as follows:

$$\text{Model 1- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi$$

$$\text{Model 2- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{unstr}(s_t)$$

$$\text{Model 3- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{str}(s_t)$$

$$\text{Model 4- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{unstr}(s_t) + f_{str}(s_t)$$

$$\text{Model 5- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi$$

$$\text{Model 6- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{unstr}(s_t)$$

$$\text{Model 7- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{str}(s_t)$$

$$\text{Model 8- logit}(p_{it}) = \beta_0 + f(age) + G'_{it}\pi + f_{unstr}(s_t) + f_{str}(s_t)$$

Model 1: This is a frequentist logistic regression model in which all the categorical variables were considered as fixed effects and assumed to have a linear effect on the response variable which include educational status, age at first sex, place of residence, marital status, contraceptive used, and STI status. ***Model 2:*** In this case, all the categorical variables listed in model 1 were assumed to have linear effects on the outcome variable, nonlinear effect of the continuous covariate age, and spatially unstructured random effect that entertain unobserved covariates that are intrinsic within the region. ***Model 3:*** All the predictor variables in this model were modeled as fixed effects and nonlinear covariate age and spatially structured random effect that cover the unobserved covariates which vary spatially within the region, described by the CAR model. ***Model 4:*** This model examines the effect of both nonlinear effects age and linear effects of categorical covariates including a convolution of spatially structured and spatially unstructured random effects, described by the CAR model and independently and identically distributed. We further consider ***Model 5–8*** as an extension of models 1–4, respectively, where CAR priors are included and that the regression coefficients π in these models are assumed to vary spatially.

2.2.2 Implementation of Conditionally Autoregressive Model in R-INLA

The spatial analysis modeling is defined under the Bayesian setup. Markov Chain Monte Carlo (MCMC) techniques have been used extensively for the computation of Bayesian inference, but there is shortcoming, which corresponds to their computational burden. To overcome the problems associated with MCMC algorithms, a new method, based on integrated nested Laplace approximations (INLA), has been proposed by Rue et al. (2009), which overcomes the problems associated with the MCMC algorithm. Hence, MCMC methods have been used to obtain estimates, but the computational time may be long if samples are highly correlated. INLA is an alternative Bayesian estimation method that computes approximations of posterior marginal distributions for latent Gaussian models, and it provides accurate estimates of the integrals through a Laplace approximation. For full discussion on INLA, see Rue et al. (2009). We also discuss the inference strategy briefly. The posterior density can be expressed as

$$\pi(x, \Psi|y) \propto \pi(\Psi)\pi(\Psi) \prod_{i=1}^{n} \psi(y_i|x_i, \Psi)$$

$$\propto \pi(\Psi)|Q(\Psi)|^{\frac{n}{2}} exp\left\{ -\frac{1}{2}x'Q(\Psi)x + \sum_{i=1}^{n} log\pi(y_i|x_i, \Psi) \right\}$$

The main goal is to estimate the desired marginal posterior distribution for the latent Gaussian model

$$\pi(x_i|y) = \int \pi(x_i|\Psi, y)\pi(\Psi|y)d\Psi \tag{7}$$

such that the posterior marginals of Ψ are estimated by

$$\pi(\Psi_i|y) = \int \pi(\Psi|y)d\Psi_{-i} \tag{8}$$

where Ψ_{-i} contains all the elements in Ψ except ψ. For easy integration of Ψ in Eqs. (7) and (8), nested approximations and numerical integration are required (Tierney & Kadane, 1986). Hence, the marginal posterior density $\pi(\Psi|y)$ of the hyper-parameters Ψ can be approximated using the Laplace approximation (Tierney & Kadane, 1986)

$$\tilde{\pi}(x_i|y) = \int \tilde{\pi}(x_i|\Psi, y)\tilde{\pi}(\Psi|y)d\Psi$$

and

$$\tilde{\pi}(\Psi_i|y) = \int \tilde{\pi}(\Psi|y)d\Psi_{-1}$$

Thus, posterior marginals can be used to compute summary statistics of interest, like posterior means, etc.

3 Results

3.1 Model Comparison Based on Deviance Information Criteria

The result presented in Table 1 shows the DICs for the eight separate models with both stationary model and spatially varying coefficients. Based on the previous studies, model with the smallest DIC gives the best fit. In the case of the model that assumes stationary model, model 2 was found to be the better model fit. In addition, the spatially varying coefficient models 5–8 are not statistically significantly different from one another as well as their corresponding stationary model. This is established from the fact that the difference in DIC is less than 3. Hence, it then implies that the covariates for HIV do not vary significantly via space. Therefore the current study presents the results based on model 8 for HIV modeling since it permits the covariates to vary spatially by the inclusion of the CAR model.

Table 1 Models comparison

	Stationary model				Spatially varying coefficients			
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
pD	14.8	23.7	23.6	23.5	18.2	18.2	18.2	18.2
DIC	3018.2	2827.6	2827.6	2827.6	2836.4	2836.6	2836.5	2836.5

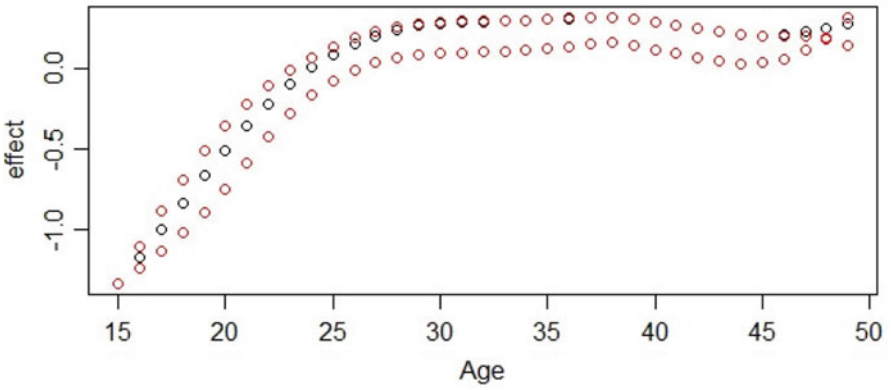


Fig. 1 Nonlinear effect of age on HIV

3.2 Nonlinear Effect of Age

The current study also examined the nonlinear association of an individual age and HIV infection as summarized in Fig. 1. It is affirm from the figure that there is a nonlinear relationship between age and HIV infection, and the assumption of linear effect would have resulted into spurious results and thereby giving inaccurate interpretations. Furthermore, the likelihood of HIV infection increases with age up to prime age of roughly 30 years and then starts to diminish afterward with increasing age.

3.3 Spatially Varying Effects

We established that the spatially varying coefficient models are better than that of stationary models as earlier acknowledged, but the models were not statistically significantly different compared to the stationary ones. Meanwhile, the maps depicted in Fig. 2 indicate that the effects of the covariates included in the models vary through space. For example, in Fig. 2, the effect of educational status on prevalence of HIV. It shows that its effect is more in Northeast and Northwest as indicated on the map. The map also affirms that age at first sex had exceptional effect in those places where education also had exceptional effects. This suggests

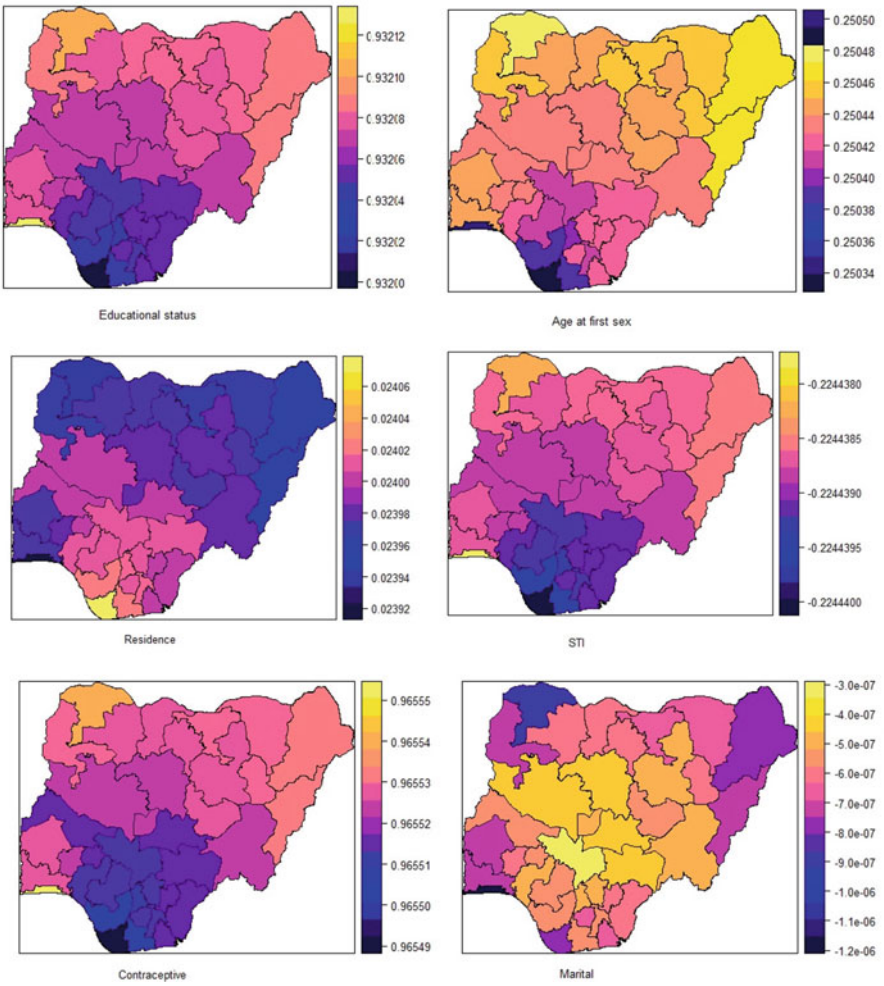


Fig. 2 The spatially varying effects of covariates utilized in modeling HIV status

that there is a correlation between educational status and age at first sex. The effect of marital status was observed to be the same across the six regions of the country, while that of place of residence was found to be more dominant in the northern region.

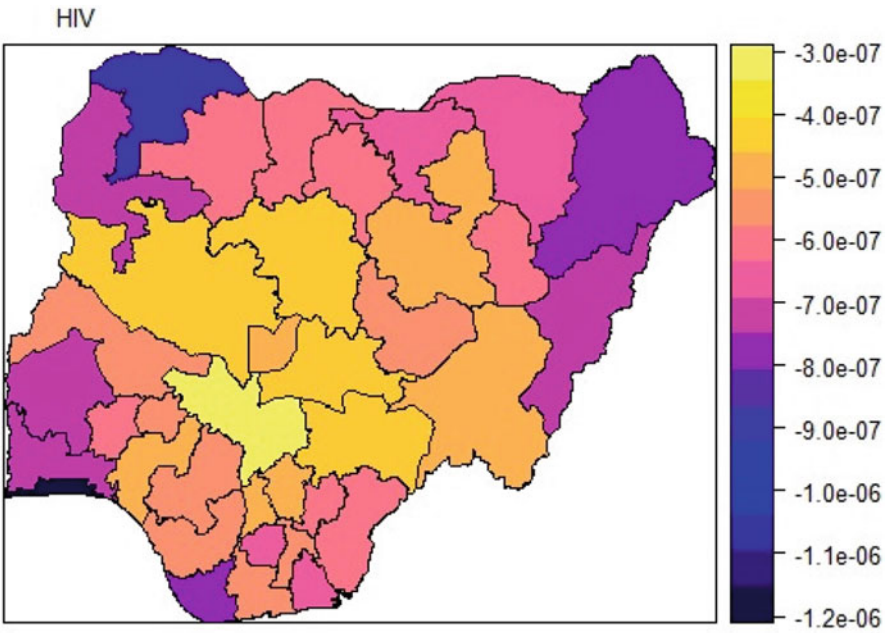


Fig. 3 Spatial effects of HIV

3.4 Spatial Effects

Based on the best fitting model among the stationary model, this study examined the spatial effects. Figure 3 presents structured spatial effects on the prevalence of HIV among sexually active women. The light blue indicates low prevalence of HIV. It was observed that the HIV prevalence varies spatially with areas in the North central and Northwest had the highest prevalence. The prevalence was lowest in the Southwest region (indicated by blue color in Fig. 3). Hence, recognizing the implications of individual covariates on each region can help informing measure to curb the prevalence of HIV.

4 Discussion

To capture the effect of age on HIV infection, this study applied Bayesian spatially varying coefficients model to estimate this nonlinearity. Various socio-demographic and sexual characteristics as well as biological risk factors of HIV were considered. Three thousand and nine hundred and seventy-five sexually active women enrolled in the demographic health survey from all the geopolitical zone of Nigeria were enrolled in the study. Based on the findings of this study, we established that the

effect of covariates on HIV infection varies spatially, and however, the spatially varying HIV model was established to be statistically insignificantly different from the two types of model examined. The use of Bayesian spatially varying model is the strength of this study as this approach is capable to reveal the effect of different covariates utilized on the prevalence of HIV across these regions.

The findings of this study showed that age has a nonlinear effect on HIV. The odds of HIV infection increase with age up to prime age of roughly 30 years and then start to diminish afterward with increasing age. Other studies have reported similar result across the world (Johnson & Way, 2006; Cohen et al., 2011) and in the continent of Africa (Ngesa et al., 2013; Michelo et al., 2006; Chimoyi & Musenge, 2014; Cogneau & Grimm, 2006). It should be noted that the spatial effects in the model act as a replacement for unobserved variables. Therefore, recognizing areas with high infection can provide more information on how the prevalence can be curbed across different regions. In addition, identifying the effects of each individual covariates on each region can also help in tackling the prevalence of HIV.

Another finding established in this study is that age at first sex had a significant effect on the prevalence of HIV in the North central and Northeast. This may be attributed to early marriages or teenage sex as most of the people residing in this part of Nigeria are largely Muslim. Implementing a community-based mobilization programs targeting on early marriages can be organized in these regions. On the other hand, this study indicates the effect of marital status on prevalence of HIV was more pronounced in the Southwest region. This result can be attributed to western education practiced in this part of Nigeria. It can also be associated with practices such as wife inheritance that is largely common in Sub-Saharan Africa (SSA), an observation that is supported by previous studies (Amornkul et al., 2009). Moreover, sexually transmitted infections were found to be a strong risk factor of HIV infection among sexually active women of Nigeria. It should be acknowledged that an incident of STIs carries duplicate chance of HIV infection and thereby increases the chance of HIV infection. This finding is supported by previous studies (Cohen, 1998) and thereby adds to the large body of scientific research showing that the role of STI in increasing HIV risk (Hankins et al., 2002; Røttingen et al., 2001).

The aim of this study was to utilize the Bayesian approach to relax the stationarity assumption in which nonlinear effects of age were captured through the random walk model of second order (Speckman & Sun, 2003). The Gaussian Markov Random Field (GMRF) was used to model the spatial and spatially unstructured random effects in the fitted model. We established that age had a nonlinear effect, while the effect of individual covariates varies across space on the prevalence of HIV. On the account of DIC, this study affirmed that the spatially varying coefficient model was fitted better than the stationary model. The findings of this suggest a significant variation of HIV prevalence across the region. These findings also indicate the region with high prevalence, and this can help policy makers as well as other public health institutions to map programmes targeting on these regions. One of the key discoveries of this chapter is that Bayesian spatially varying coefficient model used has a wider implication in the health sciences as compared to stationary

model. All the models utilized in this study were implemented with the help of software R using INLA package.

The present study has several strengths compared to previous studies. Foremost, it is a large population-based study, enhancing its generalizability compared to previous studies that has been carried out in zone that are not representative of the entire population. Also, spatially varying model to capture the nonlinearity was taken into consideration in our analytic approach, thereby resulting in unbiased estimates. However, the present study has some inherent limitations that needed to be considered when interpreting the results. One of the limitations is that a strong evidence for causality cannot be made considering cross-sectional data, but findings can contribute to knowledge and indicate where future research can be focused.

References

AIDSinfo, U. (2016). Epidemiological status.

Aitkin, M. (2010). *Statistical inference: an integrated Bayesian/likelihood approach*. Chapman and Hall/CRC.

Alexander, N. (2011). Bayesian disease mapping: Hierarchical modeling in spatial epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 512–513.

Amornkul, P. N., Vandenhoudt, H., Nasokho, P., Odhiambo, F., Mwaengo, D., Hightower, A., Buvé, A., Misore, A., Vulule, J., Vitek, C., et al. (2009). HIV prevalence and associated risk factors among individuals aged 13–34 years in rural western Kenya. *PloS One, 4*(7), e6470.

Aniekwu, N. I. (2002). Gender and human rights dimensions of HIV/aids in Nigeria. *African Journal of Reproductive Health*, 30–37.

Armstrong, A., Nagata, J. M., Vicari, M., Irvine, C., Cluver, L., Sohn, A. H., Ferguson, J., Caswell, G., Njenga, L. W., Oliveras, C., et al. (2018). A global research agenda for adolescents living with HIV. *Journal of Acquired Immune Deficiency Syndromes (1999), 78*(1), S16.

Awofala, A. A., & Ogundele, O. E. (2018). HIV epidemiology in Nigeria. *Saudi Journal of Biological Sciences, 25*(4), 697–703.

Baral, S., Beyrer, C., Muessig, K., Poteat, T., Wirtz, A. L., Decker, M. R., Sherman, S. G., & Kerrigan, D. (2012). Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet infectious diseases, 12*(7), 538–549.

Bekker, L.-G., Alleyne, G., Baral, S., Cepeda, J., Daskalakis, D., Dowdy, D., Dybul, M., Eholie, S., Esom, K., Garnett, G., et al. (2018). Advancing global health and strengthening the HIV response in the era of the sustainable development goals: the international aids society—lancet commission. *The Lancet, 392*(10144), 312–358.

Bernardinelli, L., & Montomoli, C. (1992). Empirical bayes versus fully bayesian analysis of geographical variation in disease risk. *Statistics in Medicine, 11*(8), 983–1007.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.

Beyrer, C., Baral, S. D., Collins, C., Richardson, E. T., Sullivan, P. S., Sanchez, J., Trapence, G., Katabira, E., Kazatchkine, M., Ryan, O., et al. (2016). The global response to hiv in men who have sex with men. *The Lancet, 388*(10040), 198–206.

Carlin, B. P., Gelfand, A. E., & Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Chimoyi, L. A., & Musenge, E. (2014). Spatial analysis of factors associated with hiv infection among young people in Uganda, 2011. *BMC Public Health, 14*(1), 555.

Cogneau, D., & Grimm, M. (2006). Socioeconomic status, sexual behavior, and differential aids mortality: evidence from cote D'ivoire. *Population Research and Policy Review, 25*(4), 393–407.

Cohen, M. S. (1998). Sexually transmitted diseases enhance hiv transmission: no longer a hypothesis. *The Lancet, 351*, S5–S7.

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., Hakim, J. G., Kumwenda, J., Grinsztejn, B., Pilotto, J. H., et al. (2011). Prevention of hiv-1 infection with early antiretroviral therapy. *New England Journal of Medicine, 365*(6), 493–505.

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Currie, I., & Durban, M. (2002). Flexible smoothing with p-splines: a unified approach. *Statistical Modelling, 2*(4), 333–349.

Djukpen, R. O. (2012). Mapping the hiv/aids epidemic in Nigeria using exploratory spatial data analysis. *GeoJournal, 77*(4), 555–569.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 89–102.

EO, O., Folayan, M., & Adedigba, M. (2005). Oral health-care workers and hiv infection control practices in Nigeria. *Tropical Doctor, 35*(3), 147–150.

Fahrmeir, L., & Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer Science & Business Media.

Gelfand, A. E., Kim, H.-J., Sirmans, C., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association, 98*(462), 387–396.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Group, G. H. P. W. (2003). Access to hiv prevention: closing the gap. *Foreign Affairs*, 1–42.

Hankins, C. A., Friedman, S. R., Zafar, T., & Strathdee, S. A. (2002). Transmission and prevention of hiv and sexually transmitted infections in war settings: implications for current and future armed conflicts. *Aids, 16*(17), 2245–2252.

Johnson, K., & Way, A. (2006). Risk factors for hiv infection in a national adult population: evidence from the 2003 Kenya demographic and health survey. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 42*(5), 627–636.

Lamont, J., Bajzak, K., Bouchard, C., Burnett, M., Byers, S., Cohen, T., Fisher, W., Holzapfel, S., & Senikas, V. (2012). Female sexual health consensus clinical guidelines. *Journal of Obstetrics and Gynaecology Canada, 34*(8), 769–775.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R., et al. (1999). *Disease mapping and risk assessment for public health*. New York: Wiley.

Macro, I., Commission, N. P., et al. (2014). Nigeria demographic and health survey 2013.

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis, 67*, 68–83.

Michelo, C., Sandøy, I. F., Dzekedzeke, K., Siziya, S., & Fylkesnes, K. (2006). Steep hiv prevalence declines among young people in selected Zambian communities: population-based observations (1995–2003). *BMC Public Health, 6*(1), 279.

Ngesa, O., Achia, T., & Mwambi, H. (2013). Spatial joint disease modeling and mapping with application to hiv and hsv-2. In *Annual Proceedings of the South African Statistical Association Conference*, vol. 2013 (pp. 61–68). South African Statistical Association (SASA).

Ngesa, O., Mwambi, H., & Achia, T. (2014). Bayesian spatial semi-parametric modeling of HIV variation in Kenya. *PloS One, 9*(7), e103299.

Niragire, F., Achia, T. N., Lyambabaje, A., & Ntaganira, J. (2015). Bayesian mapping of HIV infection among women of reproductive age in Rwanda. *PloS One, 10*(3), e0119944.

Pearce, M. S. (1999). Geographically weighted regression: A method for exploring spatial nonstationarity. *Stata Technical Bulletin, 8*(46).

Røttingen, J.-A., Cameron, D. W., & Garnett, G. P. (2001). A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sexually Transmitted Diseases, 28*(10), 579–597.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71*(2), 319–392.

Salaam-Blyther, T., & Kendall, A. (2012). The global fund to fight aids, tuberculosis, and malaria: Issues for congress and us contributions from fy2001 to fy2013. In *CRS Report for Congress: Prepared for Members and Committees of Congress*.

Sherman, M. (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional properties*. John Wiley & Sons.

Speckman, P. L., & Sun, D. (2003). Fully Bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika, 90*(2), 289–302.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583–639.

Team, R. C. (2016). R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing; 2015. R foundation for statistical computing.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association, 81*(393), 82–86.

Waller, L. A., Carlin, B. P., Xia, H., & Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association, 92*(438), 607–617.

WHO (2015). Consolidated guidelines on HIV testing services.

Estimating Determinants of Stage at Diagnosis of Breast Cancer Prevalence in Western Nigeria Using Bayesian Logistic Regression

Ropo Ebenezer Ogunsakin and Ding-Geng (Din) Chen

Abstract Breast cancer is the most common cancer-affecting women globally, and the stage at diagnosis remains a key factor to the final outcome. Late stage at diagnosis means a significant challenge and this is common in low middle income countries with Nigeria included. Patients with early-stage breast cancers are expected to have good survival rates. Hence, it is important to identify risk factors that predict diagnosis of early-stage breast cancers among Nigeria women. Although many studies have been carried out on the risk factors associated with breast cancer, but to the best of our knowledge little has been studied on the stage at diagnosis in the context of Nigeria. The use of advanced statistical techniques coupled with hospital-based data can enhance proper estimation of determinants of stage at diagnosis of breast cancer prevalence as well as giving appropriate explanations on the role of each determinants factors on the stage at diagnosis. In this chapter we estimated the prevalence and investigated determinants of stage at diagnosis by constructing Bayesian logistic regression model from a generalized linear modeling using socio-economic, demographic, and medical factors. It was established that age, higher educational level, being a westerner as well as choosing nursing as a career were the major factors that motivate early stage at breast cancer diagnosis in this part of Nigeria. Our findings affirmed that delays in diagnosis reflected a lack of education. This chapter suggests that further education as well as awareness of breast cancer diagnosis is required in order to increase early stage diagnosis for patients.

1 Introduction

Modern days are still challenged by diseases difficult to treat because of the lack of vaccine as well as serum. Cancer is one of the main causes of mortality

R. E. Ogunsakin (✉) · D.-G. (Din) Chen
University of Pretoria, Pretoria, South Africa
e-mail: din.chen@up.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_14

globally (Benson & Jatoi, 2012; Momenimovahed et al., 2017). For instance, in 2008, it was reported that about 8 million deaths were chronicled due to the malignant diseases, and these statistics is projected to reach 11 million by 2030 (Benson & Jatoi, 2012; Baider & Surbone, 2014). Therefore, estimating determinants of stage at diagnosis of breast cancer prevalence is vital in public health for resource allocation. Disease modeling encompasses graphic depiction of more multifaceted geographical data, which would assist in intervention, assist policy-making, and allocation of adequate resources. They are also supportive in classifying disease bunches of exact groups. With this background, this chapter focused on estimating determinants of stage at diagnosis of breast cancer prevalence in Western Nigeria, knowing fully well it is one of the significant cancer types that significantly contribute to the drain of cells in the milk-producing ducts. Breast cancer is the most common cancer among women and one of the most important causes of death among them (Ebughe et al., 2013; Chen et al., 2016). It was established that the epidemiology of breast cancer divulges various risk factors (Davis et al., 1993; Figueroa et al., 2021) and various factors contribute to its occurrence. While the disease occurs all over the globally, its incidence, mortality, and survival rates vary considerably among different parts of the world, which could be due to many factors such as population structure, lifestyle, genetic factors, and environment (Hortobagyi et al., 2005; Momenimovahed & Salehiniya, 2019). Meanwhile, changes in risk factors have resulted into an increase in the prevalence of breast cancer, which is increasing every day. In line with this, studies have revealed that various socioeconomic factors (like low levels of educational status) and demographic factors (such as age and gender) increase the menace of breast cancer. Similarly, lifestyle behaviors (such as smoking and alcohol consumption) and dietary habits also significantly influence the menace of breast cancer (Figueroa et al., 2021; Guliyeva et al., 2021).

Similarly, it is the most common cause of malignancy among women worldwide (Ebughe et al., 2013; Chen et al., 2016) and is a public health challenge among Nigeria women. In Africa and particularly Nigeria, it is the most common cancer-affecting women, and the key determinant of the final outcome is the stage at diagnosis (Ogunkorode et al., 2017; Pruitt et al., 2015). Staging is a way of describing where the cancer is located, how much the cancer has grown, and if or where it has spread. The stage of a cancer describes how much cancer is in the body. It helps determine how serious the cancer is and how best to treat it. Therefore, understanding the risk factors associated may help in curbing the menace of the disease. In western Nigeria, the disease accounted for 37% of all the newly diagnosed cancer among women attending tertiary health institution (Olugbenga et al., 2012).

In the same way, other studies affirmed that stage at diagnosis is an important prognostic factors for breast cancer (Møller et al., 2016; Kantelhardt et al., 2014). On the other hand, the stage at diagnosis of breast cancer in Nigeria has been affirmed to be late (Jedy-Agba et al., 2016, 2017). However, the presentation of breast cancer in women is done at advanced stage rather than at early symptomatic stages (Jedy-Agba et al., 2017). Therefore, knowing the relevant of early detection

and factor associated with late breast cancer presentation stage required urgent attention among the researchers. Liese Pruitt and colleagues (Pruitt et al., 2015) in a study conducted in Nigeria reported that low level of education and place of residence contribute to later stage at diagnosis. They pointed out that factors like these could bring about delay at the time of diagnosis. Other studies by Newman et al. (2015); Lipscomb et al. (2016a) mentioned that histology that is poorly differentiated may lead to later stage due to the fact that there will be excessive tumor growth rate. Meanwhile a lot of research done on the factors associated with stage at diagnosis has been reported in most other part of the world including Africa (Seneviratne et al., 2016; Akinyemiju et al., 2015b; Dickens et al., 2014) but little has been done in the context of Nigeria.

This chapter investigates how and describes the way in which stage at diagnosis varies with socio-economic, demographic, and medical factors among women seeking health care at Nigeria hospital having known that early diagnosis resulted into an excellent survival rates. More importantly, we consider the extent to which the association between stages at diagnosis is affected by a patient's age, gender, marital status, education, race, occupation, religion, site of the cancer, histological grade in order to establish which of these factors may be required as a source of curbing the menace of breast cancer in this part of Nigeria. This chapter considers two statistical approaches: the classical and Bayesian models of risk factors that affect the stage at diagnosis of patients with breast cancer. The chapter uses both the classical logistic regression model, and the Bayesian logistic regression model. The models is then applied to a dataset of patients with breast cancer in western Nigeria to establish the factors associated with stage at diagnosis. The essence and contribution of Bayesian approach is to improve the quality of the results due to the small number of observations.

2 Data and Methods

2.1 Ethical Consideration

Ethical approval for this research was obtained from Federal Medical Teaching Hospital (ERC/2016/02/25/09B). The breast cancer data were extracted from the population-based Federal Medical Teaching cancer registry. The data consist of the information recorded about patients diagnosed as having breast cancer for a period of 3 years. In these data, we defined the outcome variable as patient stage at diagnosis. Specifically, our outcome variable was stage at diagnosis whether it is early or late diagnosis. To identify the risk factors associated with stage at diagnosis, this chapter considered socio-demographic and medical factors as risk factors which includes age, marital status, educational level, religion, race, stage at diagnosis, occupation, site of the breast cancer, year of diagnosis, and tumor grade

(such as site, grade, topography, stage) and modality of treatment received: surgery, chemotherapy, hormonal therapy, radiotherapy, or combination of these.

3 Statistical Model

The data was analyzed by fitting a generalized linear model (GLM). The GLM generalizes the linear regression and relates the outcome variable to predictor variables through the link function. The statistical models, which belong to member of the GLM, include logistic regression for binary outcomes which is used to analyze the cancer data, binomial counts; binomial regression and so on. Furthermore, the GLM assumed that the variance of the response variable is a specified function of its mean. The advantage of generalized linear models is the fact that it has been extended in such a way that it can accommodate both random and mixed effects (Stroup, 2016). The logistic regression models fall into a class of generalized linear models (GLMs) technique. In general, a generalized linear model (GLM) technique, as first introduced by Nelder and Baker (1972) and modified by Fan and Gijbels (1996), provides a flexible and unified approach to analyzing both normal and non-normal data. According to McCullah and colleague (McCullagh & Nelder, 1989), the components of GLM are classified into three stages such as random component, systematic component, and link function. The link function h(.) was introduced by Nelder and Baker (1972) in order to transform the mean of the model to a linear scale. The random component is known as response variable like stage at diagnosis (in the case of our chapter) as well as its probability distribution. The fundamental idea of a GLM assumes that the underlying distribution of responses belongs to the exponential family of distributions, and a link function transformation of its expectation is modeled as a linear function of observed covariates. Meanwhile the systematic component implies the predictors (such as socio-economic, demographic, and medical factors) while the random and systematic component are linked together by the link function. Once a model is selected, there is need to estimate its parameters. In the case of GLM, the estimators of the parameters are obtained using a maximum likelihood method.

3.1 Binary Response Logistic Regression Model Formulation

The binary logistic regression is sometimes referred to as logistic regression. It is commonly applied to model data with a binary outcome (see McCullagh and Nelder (1989); Hosmer Jr. et al. (2013) for full details). The binary outcome variable assumes values one for the outcome of interest and zero for the other outcome. The predictor variables included in the model can be either continuous or categorical. Meanwhile, the probability that the value of the outcome variable is a success, given values of the predictor variables, is represented by

$$P(Y = 1|X = x) = \psi(x)$$

while the probability that it is a failure is given by

$$P(Y = 0|X = x) = 1 - \psi(x)$$

In the general linear regression model where the interest is to study the relationship between the response variable Y and the predictor variables (X_1, ..., X_k), the interest in logistic regression focuses on the relationship between the probability of the response variable being a success or otherwise failure.

In this chapter we consider logistic regression model to carry out the analysis of the data at hand. Logistic regression model is used to model the probability of occurrence of an outcome of interest, $\psi(x)$, i.e., the conditional mean of Y given x for the binomial distribution. The logistic regression model with one predictor variable in terms of the odds of the outcome of interest is given as

$$\frac{\psi}{1-\psi} = exp(\beta_0 + \beta_1 x_1) \tag{1}$$

Suppose the predictor variable X is dichotomous with values 0 and 1, hence the model in Eq. (1) can be expressed as

$$\frac{\frac{\psi(1)}{1-\psi(1)}}{\frac{\psi(0)}{1-\psi(0)}} = exp(\beta_1) \tag{2}$$

which translates that the odds ratio (OR) depends on the regression parameters β_1. On the other hand, the logistic regression model can refer to the probability of the outcome of interest and is represented as

$$\psi = \frac{exp(\beta_0 + \beta_1 x_1)}{1 + exp(\beta_0 + \beta_1 x_1)} \tag{3}$$

In this case, ψ is the expected response, E(Y|X) and β_1 is the regression coefficient. The link function, h(ψ)= log $\left\{\frac{\psi}{1-\psi}\right\}$, known as logistic function is used to transform the model in Eq. (1). The transformation also changes the range of ψ from (0 to 1) to ($-\infty$ to $+\infty$) and in turn produces a linear logistic model for the log odds of the outcome of interest and is represented as

$$log\left(\frac{\psi}{1-\psi}\right) = \beta_0 + \beta_1 x_1, \tag{4}$$

for the log odds of the outcome of interest. This model states that the log odds of the outcome of interest is linearly related with the predictor variable X_1. Also,

parameters β_0 and β_1 are the intercept and slope coefficient, respectively. Also, β_1 measures the effect of a unit change in X_1 on the log odds of the probability of the outcome of interest. It should be noted that the sign of β_1 indicates the direction of the change in ψ. When $\beta_1 > 0$, ψ increases as X_1 increases and when $\beta_1 < 0$, ψ decreases as X_1 increases.

Moreover, the logistic regression model can be extended to a situation with k predictors variables which is the focus of our chapter. In such situation, the odds of the outcome of interest can be expressed as

$$\frac{\psi}{1-\psi} = exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) \tag{5}$$

On the other hand, the logistic model may also refer directly to the probability of the outcome of interest as in the case presented in Eq. (3). This is represented as follows

$$\psi = \frac{exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)} \tag{6}$$

note that β_k is the regression parameter that shows the effect of the kth predictor variable on the log odds that Y = 1 when other predictor variables in the model are held constant. Also, $exp(\beta_k)$ is the multiplicative effect of a one unit increase in X_k , on the odds , when other predictor variables are fixed.
Using the link function, h(ψ), to transform the model in Eq. (6) gives the linear logistic model as

$$log\left(\frac{\psi}{1-\psi}\right) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \tag{7}$$

The same assumptions underlying the logistic regression model with one predictor variable are also applicable to multiple logistic regression model. The model states that the log odds of the outcome of interest is linearly related with predictor variables in the model, where β_0 is an intercept and $\beta_1, ...\beta_k$ are slope coefficients. In addition, model in Eq. (7) can also be expressed in a matrix form

$$log\left(\frac{\psi}{1-\psi}\right) = X\beta \tag{8}$$

According to Hosmer Jr. et al. (2013), X is the design matrix that includes a column of ones as the initial column to signify the constant term β_0, and thus β_1 is a vector of model parameters including the constant. In the case of logistic regression model, its specification assumes that the logit of the outcome of interest is a linear combination of the predictor variables in the model and it is clearly indicated in Eqs. (3) and (7), respectively.

The method of maximum likelihood estimate (MLE) is used to obtain the unknown parameters for logistic regression model. This approach produces values of the parameters that maximize either the likelihood or log likelihood of the parameters. Suppose we have a set of an independent observations $y_1, ..., y_n$, the log likelihood may be written as

$$\ell(\beta; y) = \sum_i \sum_i y_i x_{ij} \beta_j - \sum_{i=1}^{n} n_i log\left[1 + exp\left(\sum_{j=1}^{k} \beta_j x_{ij}\right)\right] \tag{9}$$

From Eq. (9), the likelihood depends only on y via the k linear combinations $X^{'} y$, which are the sufficient statistics for the model parameters β. The likelihood equations that emanate from differentiating the log likelihood with respect to the vector β can be represented as

$$X^{'} y = X^{'} \widehat{\mu}$$

Note that $\widehat{\mu} = n_i \widehat{\psi}_i$ .The maximum likelihood estimates satisfy the equation

$$\widehat{\beta} = (X^{'} V X)^{-1} X^{'} V$$

where V = diag $[n_i \widehat{\psi}_i (1 - \widehat{\psi}_i)]$ is the n × n diagonal matrix. Upon obtaining the maximum likelihood estimates, they can be used to make statistical inferences concerning the relationship between the response variable and predictor variables. These inferences involve assessment of the significance of predictor variables in the logistic regression model. The assessment is done by formulating and testing a statistical hypothesis that the predictor variables in the model are significantly related to the response variable.

The Wald test statistic is used to test for the significance of each coefficient β in the model.

$$Z = \frac{\widehat{\beta}}{SE(\widehat{\beta})}$$

where $\widehat{\beta}$ is the maximum likelihood estimate of β and SE $(\widehat{\beta})$ is the standard error of the estimate. In the same vein ,in the case of multiple logistic regression, the Wald test statistic is

$$V = \widehat{\beta}' [cov(\widehat{\beta})]^{-1} \widehat{\beta}$$

This has a chi-square distribution with k degrees of freedom, where k is the rank of the covariance matrix. In addition, Eq. (8) can be re-written as

$$\psi(x) = h^{-1}\left(x_i^{'}\beta\right)$$

$$= \frac{exp\left(x_i^{'}\beta\right)}{1 + exp\left(x_i^{'}\beta\right)}$$

Hence, the likelihood function can be written as

$$Likelihood = \prod_{n=1}^{k}\left[\left(\frac{exp(x_i^{'}\beta)}{1 + exp(x_i^{'}\beta)}\right)^{y_i}\left(1 - \frac{exp(x_i^{'}\beta)}{1 + exp(x_i^{'}\beta)}\right)^{1-y_i}\right] \tag{10}$$

Under a Bayesian paradigm, the prior distributions are the respective distributions of the set of parameters β_0, ..., β_k on which the choice for priors depends on available information. One of the common priors is of the form given below

$$\beta_i \sim N(\mu_i, \sigma_i^2), \quad \sigma_i^2 \sim Inv - \chi^2(\nu_i, s_i^2) \tag{11}$$

where μ_i is sometimes taken to be zero and σ often chosen to be large so that the prior can be non-informative, ν and s denote degrees of freedom and scale for the t-prior distributions, respectively.

The posterior distribution is obtained by combining the full likelihood function in Eq. (10) and the prior in Eq. (11) to obtain

$$\begin{aligned} posterior = \prod_{n=1}^{k}\left[\left(\frac{exp(x_i^{'}\beta)}{1 + exp(x_i^{'}\beta)}\right)^{y_i}\left(1 - \frac{exp(x_i^{'}\beta)}{1 + exp(x_i^{'}\beta)}\right)^{1-y_i}\right] \\ \cdot \prod_{i=0}^{k}\frac{1}{\sqrt{2\pi\sigma_i}}exp\left[-\frac{1}{2}\left(\frac{\beta_i - \mu_i}{\sigma_i}\right)^2\right] \end{aligned} \tag{12}$$

Now, we found that the expression Eq. (12) is a complex function of the parameters, and numerical approaches are required in order to obtain the marginal posterior distribution for each of the model parameters. Approximations can be obtained via numerical integration (Naylor & Smith, 1982). This chapter used Markov Chain Monte Carlo (MCMC) technique to simulation of the random numbers as a result of the complexity of the posterior.

3.1.1 Computations and Implementations

Analyses of this research are based on the hospital-based data. All the analyses were carried out in R packages, Team (2013) (Team et al., 2013). In particular, the package **arm** (applied regression and multilevel modelling) by De Leeuw et al. (2008)) was used to compute the Bayesian logistic regression. The natural non-informative prior density is uniform on s_i^2. In the model, the posterior distribution for the parameters β and σ can be simulated using Gibbs sampler. On the other hand, updating the vector β given σ with normal regression and updating the vector σ from the independent inverse-χ^2 conditional posterior distributions given β. If the coefficients β_i have independent t prior distributions along centers μ_i and scales s_i, the iterative weighted least squares can be employed to estimate the coefficients using an EM algorithm. The idea is to express the t prior distribution for each coefficient β_i as a mixture of normal with unknown scale. As a requirement of the Bayesian approach several diagnosis test were perform to answer convergence of the Markov chain Monte Carlo algorithm and the true reflection of the posterior distribution. Due to the binary nature of the response variable, a generalized linear model (GLM) with binary outcome and logistic link were performed both using classical techniques as well as Bayesian techniques.

4 Results

The main purpose of this paper is to establish significant predictors of stage of BC diagnosis, using classical approach and Bayesian approach. In order to achieve this objective, we set up a generalized linear model for the two approaches. The predictors included in the model are socio-demographic and medical factors. From the descriptive results, we found that the mean age at BC diagnosis was 42.2 years ($\pm$16.6 years). The majority (57%) were recruited from a university teaching hospital. In all, 121(49.4%) BC patients were found to have grade II tumor at the time of diagnosis. The graphical representation of the stage at BC diagnosis was constructed in order to complement the result of the descriptive statistics.

The classical and Bayesian (logistic) generalized linear models were fitted using the same covariates. The outcome variable is early BC stages or late BC stages(being the reference category). In order to establish the significant level for the predictors, we used 95% confidence interval for classical approach and 95% credible interval for Bayesian approach. In case these interval contains zero, it then means that the parameter (estimate of the beta) is not significant.

This section gives details of how the logistic regression was used to construct a classical model for stage at diagnosis of breast cancer. The data analysis was done using R software. The objective of this chapter is to identify the individual characteristics that could be associated with the stage at diagnosis among breast cancer patients. The predictors included are age, educational status, marital status, occupation, race, religion, tumor grade, and topography.

Table 1 Estimates and odds ratios showing association between early stage at diagnosis of BC and predictor variables

	Estimate	OR	95%CI		P-value
			Lower	Upper	
Intercept	4.279	0.013	0.0004	0.205	0.004
Age group (ref: 20–34)					
35–49	1.364	3.912	1.211	14.914	0.031
50–64	0.470	1.601	0.288	8.852	0.583
65+	1.198	3.312	0.419	36.362	0.278
Education (ref: primary)					
Secondary	2.210	9.116	0.993	226.533	0.086
Tertiary	3.382	29.430	2.659	815.401	0.014
Marital status (ref: single)					
Married	0.421	1.524	0.199	18.458	0.708
Religion (ref: Muslim)					
Christian	0.995	2.705	0.677	10.528	0.148
Race (ref: igbo/efik)					
Yoruba	0.971	2.640	1.106	6.204	0.026
Occupation (ref: civil servant)					
Nurse	3.368	29.020	2.816	797.549	0.013
Retired	0.668	1.950	0.281	17.597	0.511
Self-employed	1.165	3.206	0.976	11.254	0.058
Tumor grade (ref: poorly differentiated)					
Moderately differentiated	−0.009	0.991	0.415	2.304	0.984
Well differentiated	−0.044	0.957	0.313	3.019	0.938
Topography (ref: over lesion)					
Lower inner	15.643	$6.21e^{6}$	$1.198e^{-6}$	$1.365e^{44}$	0.985
Breast NOS	−0.397	0.672	0.261	1.826	0.421

Table 1 presents the covariates that were associated with early stage at breast cancer diagnosis in the classical model. Findings from this model affirm that four major variables out of all the variables included in the model were significant predictors associated with the stage at diagnosis at 95% significant level or above. Based on the result, patients in the age category of 35–49 years are 61% (OR=3.912) more likely to be associated with early stage at diagnosis compared to age category 20–34 years. There are many reasons that could be associated with this kind of finding. Also, results reveal that patient with higher educational level are 71% (OR=29.4) more likely to be associated with early stage compared to those with primary school education. Similarly, those who are medical practitioner (nurse) are more likely to be associated with early stage at diagnosis compared to the civil servants. Furthermore, compared to northerner, patients who are from western region were 2.60 times more likely to have early stage at diagnosis. Additionally, results about marital status affirmed that patients who have married are 85%

(OR=1.50) more likely to be associated with early stage at diagnosis but the finding is not significant.

4.1 Bayesian Logistic Regression

The same covariates used in classical model are included in the Bayesian model. The results obtained in Bayesian model are given in Table 2. In terms of statistical significant, the result of Bayesian model is very similar to that of classical model but the two techniques are difficult to compare due to the difference in tools used for decision making. The most important aspect of Bayesian statistics is that the credible interval is very different from the confidence interval used for classical statistics because credible interval is more robust than that of confidence interval. The results in Table 2 affirm that patient aged 35 to 49 years are 1.17 times more likely of having early stage at diagnosis compared to age group 20 to 34

Table 2 Bayesian logistic regression model estimates for early stage at BC diagnosis

		95%Cred.I	
	Estimate	Lower	Upper
Intercept	−0.024	−0.371	0.322
Age group (ref: 20–34)			
35–49	0.158	0.017	0.299
50–64	0.0758	−0.109	0.257
65+	0.142	−0.071	0.355
Education (ref: primary)			
Secondary	0.470	0.163	0.776
Tertiary	0.335	0.043	0.629
Marital status (ref: single)			
Married	0.036	−0.197	0.267
Religion (ref: Muslim)			
Christian	0.153	−0.051	0.355
Race (ref: igbo/efik)			
Yoruba	0.146	0.022	0.269
Occupation (ref: civil servant)			
Nurse	−0.305	0.070	0.539
Retired	0.065	−0.151	0.282
Self-employed	0.103	−0.035	0.239
Tumor grade (ref: poorly differentiated)			
Moderately differentiated	0.013	−0.093	0.121
Well differentiated	0.018	−0.123	0.159
Topography (ref: over lesion)			
Lower inner	0.115	−0.230	0.461
Breast NOS	−0.035	−0.111	0.099

years. With reference to primary education, the results show that patient with tertiary education are more likely to be diagnosed at early stage (OR=1.690, 95% BCI=0.163, 0.776) while those with secondary school education are 1.4 times more likely to be associated with early diagnosis of BC. Regarding the occupational status, the chances of being at early stage of diagnosis are more (OR=1.357, 95% BCI=0.070, 0.539) among those who are medical practitioners compared to the civil servants. For the religion factor, with reference to those who practiced Islamic religion, it is over one times more likely (OR=1.165, 95% BCI=−0.051, 0.355) for those who are Christian to have early stage at diagnosis. For marital status, women who are married are 1.037 times more likely to be diagnosed at early stages than those who are single (OR=1.037, 95% BCI=−0.197, 0.267). This means that the only significant effect here was that single women were generally at higher risk to advanced stage than married women.

For the MCMC convergence, diagnostic tests are used to assess the convergence of the Markov chain. The posterior distribution was obtained after two thousand iterations performed gradually and assessing convergence at every stage. Figure 1 gives the trace plots for a few of the parameters of the posterior distribution obtained by the MCMC algorithm. All the trace plots do not display any significant upward or downward trend along the iterations and the density plots also show almost symmetrical distributions. In particular, the trace plots exhibit the so-called thick pen as explained by Gelfand et al. (1990). Hence, this is indicative of insignificant deviations from stationarity and the MCMC algorithm can be considered to have converged.

Figure 2 gives the graphical representation of Geweke plots for some selected parameters used in the model. As a rule of thumb, a significant proportion of Z-

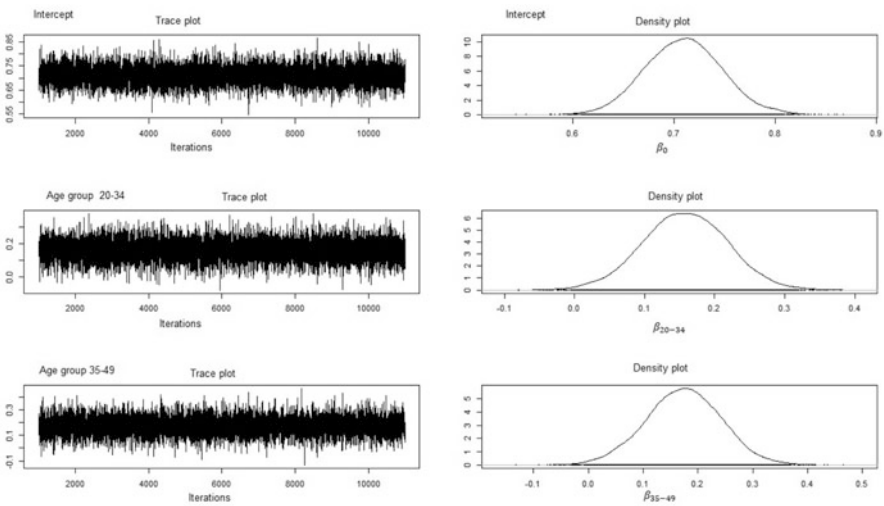


Fig. 1 Trace plots and density plots for the first three coefficients from the posterior distribution

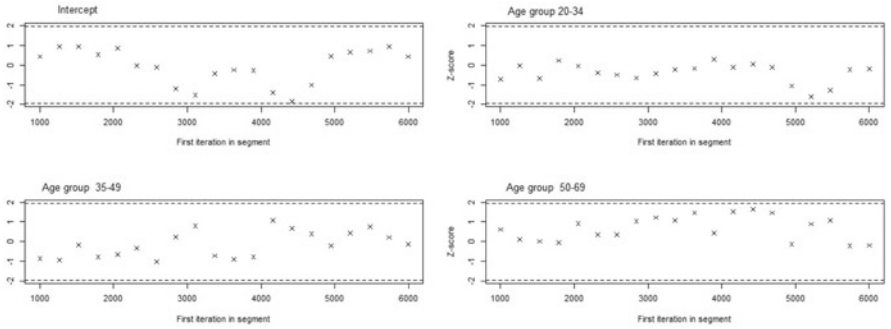


Fig. 2 Geweke plots for the first four coefficients from the posterior distribution

scores outside the two-standard deviation bands is an indication of a chain that has not converged by iteration k. The results in Fig. 2 show that all the Z-scores fall within the two-standard deviation bands for the parameters age groups 20–34, 35–49, and 50–69. This is a strong indication of a chain that has converged by iteration. Also, we considered the Heidelberger–Welch diagnostic test. The results of a test with $\epsilon = 0.1$ show that most of the parameters have passed the stationarity test. All the parameters passed the half-width test meaning that the chain was run sufficiently long.

5 Discussion

In an effort to improve the understanding of the breast cancer menace in Nigeria, this paper has reported on breast cancer prevalence in western Nigeria and assessed key variables for their association with stage at diagnosis. The objective of this paper was to identify key socio-demographic, and medical as risk factors for stage at diagnosis of BC in western Nigeria. This chapter utilized classical and Bayesian techniques to analyze determinants and risk factors associated with early stage at diagnosis of BC. The chapter developed and used Bayesian logistic regression models to help assess factors associated with early stage diagnosis of BC. Our variables were categorized as key socio-demographic, and medical factors were analyzed by using logistic regression model contained in software R. Our chapter identified age group 35–49 years, secondary and tertiary educational level, ethnicity and choosing nursing as career as independent associates of early stage at BC diagnosis in this part of Nigeria. On the other hand, marital status, tumor grade, and topography were not found to be associated with stage at BC diagnosis. This is in contrast with a what is obtainable in South Africa previous studies where it was reported that late stage were more likely to be estrogen but showed no association with human epidermal growth factor receptor 2 (Jedy-Agba et al., 2017). The mean age at breast cancer diagnosis obtained in this chapter was similar to what is reported by previous studies in Nigeria (Ezeome, 2010; Anyanwu et al.,

2011) and other part of the world (Galukande et al., 2013; Mabula et al., 2012). In this chapter we found an association between age at diagnosis and stage. Previous studies have also observed an association between younger age at diagnosis and later stage (de Souza Abrahão et al., 2015) while other reported no association (Lipscomb et al., 2016b).

In this chapter we found that level of education was a strong factor of stage at diagnosis. Women with tertiary educational level had significantly higher odds of early stage than their counterpart with secondary educational level. This finding is supported by previous studies and, hence, adds to the large body of research indicating that level of education motivates early breast diagnosis (Jedy-Agba et al., 2017; de Souza Abrahão et al., 2015; Olugbenga et al., 2012; Pace et al., 2015). Other studies from Nigeria (Pruitt et al., 2015; Abimbola, 2010) have observed that low level of educational status was associated with early stage of BC diagnosis. We can also attribute the significance of tertiary education which resulted to early stage diagnosis to observation of appropriate caution to life management as well as increased awareness and curiosity to prolong life among this population. Other studies established that there is a correlation between breast cancer knowledge and the level of education (Soyer et al., 2007) and absence of crucial knowledge can have a detrimental effect on the attitudes of women in the adoption of early stage BC diagnosis. Although our chapter did not examine the risk factors for breast cancer association with its sub molecular types, a recent study conducted by Akinyemiju et al. (2015a); Ogunsakin and Siaka (2017) evaluated the association between SES and breast cancer subtypes using a valid measure of SES and the Surveillance, Epidemiology and End Results (SEER) database. Socio-economic status based on measures of income, occupational class, education, and house value were categorized into quintiles and explored. Their findings showed that a positive association between SES and breast cancer incidence is primarily driven by hormone receptor positive lesion.

Another finding from this chapter is that women who were from Christianity religion were diagnosed at early stages compared to Islamic women. These findings support the previous studies that have reported less BC early detection practices among the Islamic women. The implication of our finding might be these women probably belong to a higher socio-economic class than the Islamic women and, hence, have financial strength to seek medical care as at when due. Additionally, being a nurse was found to be associated with the early stage BC diagnosis. This finding agrees with the previous studies in Nigeria that education and employment in professional jobs significantly influenced knowledge of breast cancer (Jebbin & Adotey, 2004; Okobia et al., 2006; Odusanya & Tayo, 2001; Odusanya, 2001; Akhigbe & Omuemu, 2009; Clegg et al., 2009; Byers et al., 2008; Barry et al., 2012). We also found from this chapter that race is an important factor for early stage at diagnosis in this part of Nigeria. Our finding indicated that being a western woman could have resulted to early stage in diagnosis of breast cancer. Obviously, the majority of people in this part of Nigeria are mostly westerner but are more knowledgeable compared to other tribes. Hence, this finding might be related to advancement in education, technology, and exposure to western diets and culture of

the people from these part of Nigeria (Olugbenga et al., 2012; Lantz et al., 2006). This finding contrast the previous studies that reported that in all age groups, black race was associated with being diagnosed beyond stage I (Iqbal et al., 2015). In another development, it was established that there is variation in early stage BC by region (Wang et al., 2008) and distance to health care facilities may increase the likelihood of stage at diagnosis as reported by Jedy-Agba et al. (2017). Contrary to what is obtainable in our chapter, a study conducted in other part of the world indicated that there are race/ethnicity differences in breast cancer stage at diagnosis, with African American women significantly more likely than white women to have a late-stage diagnosis (Mandelblatt et al., 1991; Hunter et al., 1993; Batina et al., 2013).

Furthermore, previous studies also affirmed that tumor grade and later stage at BC diagnosis were associated (McCormack et al., 2013). Meanwhile our chapter did not investigate the factors associated with late stage at diagnosis but in a study conducted in Nigeria by Jeddy and colleagues (Jedy-Agba et al., 2017) reported that there is no indication of association between tumor grade and late stage at diagnosis. In the case of our chapter we observed no evidence of association between the tumor grade and early stage. Other studies involving women from New Zealand (Seneviratne et al., 2016) have reported poorly differentiated tumor in women with late stage. A further study by Limpscomb and colleagues (Lipscomb et al., 2016a) similarly revealed that advanced stage of BC was positively related to poorly differentiated tumor grade. There is a necessary call for awareness to enlighten people particularly among the women of low socio-economic status in western Nigeria. This will aid early BC diagnosis and bring about the intervention to reduce the menace of the disease in Nigeria.

We computed a classical and Bayesian logistic regression model from a GLM perspective for stage at BC diagnosis on socio-demographic and medical variables. Non-informative prior probability distributions for the socio-demographic and medical factors variables were used in the building of the model. This chapter suggests that stage at BC diagnosis is related to an individual's age, educational status, ethnicity, and choosing nursing as career. As compared to the single/separated, the married individuals are more likely to be associated with early stage diagnosis. However, the chapter also suggests that marital status, tumor grade, and topography are not statistically significant associated with the stage at diagnosis of BC. Another key finding of this paper is that Bayesian technique is more robust and helps in selecting the more statistical significant factors associated with early stage diagnosis of BC in western Nigeria. We recommend that further education as well as awareness of breast cancer diagnosis is required in order to increase early stage diagnosis among breast cancer patients. In addition, more investigation on breast cancer research is required to characterize delays and associated factors with diagnostic delays in western Nigeria as such investigation is necessary for breast cancer control.

6 Strengths and Limitations of the Chapter

One of the key strength of this paper is the fact that it makes use of appropriate statistical techniques coupled with the use of advanced statistical software in estimating determinants of stage of disease at diagnosis. Besides using patients from some selected tertiary hospitals, the included patients represented a significant geographic range in western Nigeria. The two major religions in the area were also present. However, information regarding the wealth index of the patients was not captured. This could have shown the financial capability of the patient if the breast cancer was detected earlier whether they can afford the required treatment. Another drawback is the fact that the chapter used secondary data which limited the control of the analyst over the data collection. Meanwhile, this limitation does not relegate that the hospital-based record data is not good for research. Taken together, these studies have significantly contributed to our understanding of risk factors associated with an early stage diagnosis of breast cancer in western Nigeria. Although, there are certain methodological considerations, not explored in the current chapter, with potentially important implications for the specification of models and interpretation of findings.

7 Implications of the Chapter

In the field of scientific and health related research, find the statistical models that incorporate prior known information about the unknown model parameters vital. This is useful in studies where replicative experimental investigations are not possible. The most useful part of Bayesian statistical paradigm is in combining prior knowledge about model parameters with the appropriate likelihood of the observed data to obtain a posterior distribution. Under the Bayesian framework, likelihood based approaches are often used for parameter estimation whilst statistical inference is carried out based on the posterior distribution. The computer intensive simulation-based algorithms like Markov chain Monte Carlo (MCMC) methods are then used to draw samples from the posterior distribution to be used for the statistical inference purposes. In addition, to assess convergence which is a requirement of the Markov chain, diagnostics in the form of trace plots and Geweke plots as well as Heidelberg test for stationarity were employed.

There has been a host of prior knowledge about breast cancer stages at diagnosis that can be combined with the likelihood of the observed data to enhance explaining the variation of stages at diagnosis. The use of hospital population-based data also facilitates linking of breast cancer stages at diagnosis to socio-demographic and medical factors of the respondents. Therefore, this chapter fitted a Bayesian logistic regression model from a generalized linear modelling (GLM) perspective for stages at diagnosis on socio-demographic and medical factors using a hospital-based data. The unknown model parameters employed a non-informative t-family Cauchy prior

distribution. Finally, it was affirmed that stages at diagnosis in western Nigeria are dependent on one's age, educational level, and choosing nursing as a career.

References

Abimbola, O. O. (2010). Assessment of women's risk factors for breast cancer and predictors of the practice of breast examination in two rural areas near Ibadan, Nigeria. *Cancer Epidemiology, 34*(4), 425–428.

Akhigbe, A. O., & Omuemu, V. O. (2009). Knowledge, attitudes and practice of breast cancer screening among female health workers in a Nigerian urban city. *BMC Cancer, 9*(1), 203.

Akinyemiju, T. F., Pisu, M., Waterbor, J. W., & Altekruse, S. F. (2015a). Socioeconomic status and incidence of breast cancer by hormone receptor subtype. *SpringerPlus, 4*(1), 508.

Akinyemiju, T. F., Vin-Raviv, N., Chavez-Yenter, D., Zhao, X., & Budhwani, H. (2015b). Race/ethnicity and socio-economic differences in breast cancer surgery outcomes. *Cancer Epidemiology, 39*(5), 745–751.

Anyanwu, S. N., Egwuonwu, O. A., & Ihekwoaba, E. C. (2011). Acceptance and adherence to treatment among breast cancer patients in eastern Nigeria. *The Breast, 20*, S51–S53.

Baider, L., & Surbone, A. (2014). Universality of aging: family caregivers for elderly cancer patients. *Frontiers in Psychology, 5*, 744.

Barry, J., Breen, N., & Barrett, M. (2012). Significance of increasing poverty levels for determining late-stage breast cancer diagnosis in 1990 and 2000. *Journal of Urban Health, 89*(4), 614–627.

Batina, N. G., Trentham-Dietz, A., Gangnon, R. E., Sprague, B. L., Rosenberg, M. A., Stout, N. K., Fryback, D. G., & Alagoz, O. (2013). Variation in tumor natural history contributes to racial disparities in breast cancer stage at diagnosis. *Breast Cancer Research and Treatment, 138*(2), 519–528.

Benson, J. R., & Jatoi, I. (2012). The global breast cancer burden. *Future Oncology, 8*(6), 697–702.

Byers, T. E., Wolf, H. J., Bauer, K. R., Bolick-Aldrich, S., Chen, V. W., Finch, J. L., Fulton, J. P., Schymura, M. J., Shen, T., Van Heest, S., et al. (2008). The impact of socioeconomic status on survival after cancer in the united states. *Cancer, 113*(3), 582–591.

Chen, H.-l., Zhou, M.-q., Tian, W., Meng, K.-x., & He, H.-f. (2016). Effect of age on breast cancer patient prognoses: A population-based study using the seer 18 database. *PloS One, 11*(10), e0165409.

Clegg, L. X., Reichman, M. E., Miller, B. A., Hankey, B. F., Singh, G. K., Lin, Y. D., Goodman, M. T., Lynch, C. F., Schwartz, S. M., Chen, V. W., et al. (2009). Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National longitudinal mortality study. *Cancer Causes & Control, 20*(4), 417–435.

Davis, D. L., Bradlow, H. L., Wolff, M., Woodruff, T., Hoel, D. G., & Anton-Culver, H. (1993). Medical hypothesis: xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives, 101*(5), 372–377.

De Leeuw, J., Meijer, E., & Goldstein, H. (2008). Handbook of multilevel analysis.

de Souza Abrahão, K., Bergmann, A., de Aguiar, S. S., & Thuler, L. C. S. (2015). Determinants of advanced stage presentation of breast cancer in 87,969 Brazilian women. *Maturitas, 82*(4), 365–370.

Dickens, C., Joffe, M., Jacobson, J., Venter, F., Schüz, J., Cubasch, H., & McCormack, V. (2014). Stage at breast cancer diagnosis and distance from diagnostic hospital in a peri-urban setting: a south African public hospital case series of over 1000 women. *International Journal of Cancer, 135*(9), 2173–2182.

Ebughe, G., Ekanem, I., Omoronyia, O., Nnoli, M., Nwagbara, V., Udosen, J., Umoh, M., and Ugbem, T. (2013). Age specific incidence of breast cancer in Calabar, Nigeria.

Ezeome, E. R. (2010). Delays in presentation and treatment of breast cancer in Enugu, Nigeria. *Nigerian Journal of Clinical Practice, 13*(3).

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, vol. 66. CRC Press.

Figueroa, J. D., Gierach, G. L., Duggan, M. A., Fan, S., Pfeiffer, R. M., Wang, Y., Falk, R. T., Loudig, O., Abubakar, M., Ginsberg, M., et al. (2021). Risk factors for breast cancer development by tumor characteristics among women with benign breast disease. *Breast Cancer Research, 23*(1), 1–12.

Galukande, M., Wabinga, H., Mirembe, F., Karamagi, C., and Asea, A. (2013). Difference in risk factors for breast cancer by ER status in an indigenous African population. *ISRN Oncology*, 2013.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association, 85*(412), 972–985.

Guliyeva, G., Huayllani, M. T., Boczar, D., Avila, F. R., Lu, X., & Forte, A. J. (2021). Age as a risk factor for breast cancer-related lymphedema: a systematic review. *Journal of Cancer Survivorship*, 1–8.

Hortobagyi, G. N., de la Garza Salazar, J., Pritchard, K., Amadori, D., Haidinger, R., Hudis, C. A., Khaled, H., Liu, M.-C., Martin, M., Namer, M., et al. (2005). The global breast cancer burden: variations in epidemiology and survival. *Clinical Breast Cancer, 6*(5), 391–401.

Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*, vol.398. John Wiley & Sons.

Hunter, C. P., Redmond, C. K., Chen, V. W., Austin, D. F., Greenberg, R. S., Correa, P., Muss, H. B., Forman, M. R., Wesley, M. N., Blacklow, R. S., et al. (1993). Breast cancer: factors associated with stage at diagnosis in black and white women. *JNCI: Journal of the National Cancer Institute, 85*(14), 1129–1137.

Iqbal, J., Ginsburg, O., Rochon, P. A., Sun, P., and Narod, S. A. (2015). Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the united states. *Jama, 313*(2), 165–173.

Jebbin, N. J., & Adotey, J. M. (2004). Attitudes to, knowledge and practice of breast self-examination (BSE) in port Harcourt. *Nigerian Journal of Medicine: Journal of the National Association of Resident Doctors of Nigeria, 13*(2), 166–170.

Jedy-Agba, E., McCormack, V., Adebamowo, C., & dos Santos-Silva, I. (2016). Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. *The Lancet Global Health, 4*(12), e923–e935.

Jedy-Agba, E., McCormack, V., Olaomi, O., Badejo, W., Yilkudi, M., Yawe, T., Ezeome, E., Salu, I., Miner, E., Anosike, I., et al. (2017). Determinants of stage at diagnosis of breast cancer in Nigerian women: sociodemographic, breast cancer awareness, health care access and clinical factors. *Cancer Causes & Control, 28*(7), 685–697.

Kantelhardt, E. J., Zerche, P., Mathewos, A., Trocchi, P., Addissie, A., Aynalem, A., Wondemagegnehu, T., Ersumo, T., Reeler, A., Yonas, B., et al. (2014). Breast cancer survival in Ethiopia: a cohort study of 1070 women. *International Journal of Cancer, 135*(3), 702–709.

Lantz, P. M., Mujahid, M., Schwartz, K., Janz, N. K., Fagerlin, A., Salem, B., Liu, L., Deapen, D., & Katz, S. J. (2006). The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. *American Journal of Public Health, 96*(12), 2173–2178.

Lipscomb, J., Fleming, S. T., Trentham-Dietz, A., Kimmick, G., Wu, X.-C., Morris, C. R., Zhang, K., Smith, R. A., Anderson, R. T., & Sabatino, S. A. (2016a). What predicts an advanced-stage diagnosis of breast cancer? sorting out the influence of method of detection, access to care and biological factors. *Cancer Epidemiology and Prevention Biomarkers*, cebp–0225.

Lipscomb, J., Fleming, S. T., Trentham-Dietz, A., et al. (2016b). Centers for disease control and prevention national program of cancer registries patterns of care study group. what predicts an advanced-stage diagnosis of breast cancer? sorting out the influence of method of detection, access to care, and biologic factors. *Cancer Epidemiol Biomarkers & Prevention, 25*(4), 613–623.

Mabula, J. B., Mchembe, M. D., Chalya, P. L., Giiti, G., Chandika, A. B., Rambau, P. F., Masalu, N., and Gilyoma, J. M. (2012). Stage at diagnosis, clinicopathological and treatment patterns of breast cancer at Bugando medical centre in north-western Tanzania. *Tanzania Journal of Health Research, 14*(4).

Mandelblatt, J., Andrews, H., Kerner, J., Zauber, A., & Burnett, W. (1991). Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *American Journal of Public Health, 81*(5), 646–649.

McCormack, V. A., Joffe, M., van den Berg, E., Broeze, N., dos Santos Silva, I., Romieu, I., Jacobson, J. S., Neugut, A. I., Schüz, J., & Cubasch, H. (2013). Breast cancer receptor status and stage at diagnosis in over 1,200 consecutive public hospital patients in Soweto, South Africa: a case series. *Breast Cancer Research, 15*(5), R84.

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models, no. 37 in monograph on statistics and applied probability.

Møller, H., Henson, K., Lüchtenborg, M., Broggio, J., Charman, J., Coupland, V. H., Davies, E., Jack, R. H., Sullivan, R., Vedsted, P., et al. (2016). Short-term breast cancer survival in relation to ethnicity, stage, grade and receptor status: national cohort study in England. *British Journal of Cancer, 115*(11), 1408.

Momenimovahed, Z., Ghoncheh, M., Pakzad, R., Hasanpour, H., & Salehiniya, H. (2017). Incidence and mortality of uterine cancer and relationship with human development index in the world. *Cukurova Medical Journal, 42*(2), 233–240.

Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy, 11*, 151.

Naylor, J. C., & Smith, A. F. (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, 214–225.

Nelder, J. A., & Baker, R. J. (1972). *Generalized linear models*. Wiley Online Library.

Newman, L. A., Reis-Filho, J. S., Morrow, M., Carey, L. A., & King, T. A. (2015). The 2014 society of surgical oncology Susan G. Komen for the cure symposium: Triple-negative breast cancer. *Annals of Surgical Oncology, 22*(3), 874–882.

Odusanya, O. O. (2001). Breast cancer: knowledge, attitudes, and practices of female schoolteachers in Lagos, Nigeria. *The Breast Journal, 7*(3), 171–175.

Odusanya, O. O., & Tayo, O. O. (2001). Breast cancer knowledge, attitudes and practice among nurses in Lagos, Nigeria. *Acta Oncologica, 40*(7), 844–848.

Ogunkorode, A., Holtslander, L., Anonson, J., & Maree, J. (2017). Promoting early detection of breast cancer and care strategies for Nigeria. *African Journal of Reproductive Health, 21*(2), 18–25.

Ogunsakin, R. E., & Siaka, L. (2017). Bayesian inference on malignant breast cancer in Nigeria: A diagnosis of MCMC convergence. *Asian Pacific Journal of Cancer Prevention: APJCP, 18*(10), 2709.

Okobia, M. N., Bunker, C. H., Okonofua, F. E., & Osime, U. (2006). Knowledge, attitude and practice of Nigerian women towards breast cancer: a cross-sectional study. *World Journal of Surgical Oncology, 4*(1), 11.

Olugbenga, A. M., Olanrewaju, M. J., & Kayode, O. M. (2012). Profile of cancer patients attending tertiary health institutions in southwestern Nigeria. *Asian Journal of Pharmaceutical and Clinical Research, 5*(1), 34–37.

Pace, L. E., Mpunga, T., Hategekimana, V., Dusengimana, J.-M. V., Habineza, H., Bigirimana, J. B., Mutumbira, C., Mpanumusingo, E., Ngiruwera, J. P., Tapela, N., et al. (2015). Delays in breast cancer presentation and diagnosis at two rural cancer referral centers in Rwanda. *The Oncologist, 20*(7), 780–788.

Pruitt, L., Mumuni, T., Raikhel, E., Ademola, A., Ogundiran, T., Adenipekun, A., Morhason-Bello, I., Ojengbede, O. A., & Olopade, O. I. (2015). Social barriers to diagnosis and treatment of breast cancer in patients presenting at a teaching hospital in Ibadan, Nigeria. *Global Public Health, 10*(3), 331–344.

Seneviratne, S., Lawrenson, R., Harvey, V., Ramsaroop, R., Elwood, M., Scott, N., Sarfati, D., & Campbell, I. (2016). Stage of breast cancer at diagnosis in New Zealand: impacts of socio-demographic factors, breast cancer screening and biology. *BMC Cancer, 16*(1), 129.

Soyer, M. T., Ciceklioglu, M., & Ceber, E. (2007). Breast cancer awareness and practice of breast self examination among primary health care nurses: influencing factors and effects of an in-service education. *Journal of Clinical Nursing, 16*(4), 707–715.

Stroup, W. W. (2016). *Generalized linear mixed models: modern concepts, methods and applications*. CRC Press.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Wang, F., McLafferty, S., Escamilla, V., & Luo, L. (2008). Late-stage breast cancer diagnosis and health care access in Illinois. *The Professional Geographer, 60*(1), 54–69.

Part V
Statistical Applications

Identifying Outlying and Influential Clusters in Multivariate Survival Data Models

Tsirizani M. Kaombe and Samuel O. M. Manda

Abstract In regression analysis, diagnostic statistics serve to assess the quality of fit of the model to data and investigate if there are observations that are not well represented by the model. Additionally, there could be outlier observations in the sense that they deviate from the pattern of the other data points being modelled, or influential observations that, if removed from the dataset, could impact the slope of the fitted regression model. These two types of unusual data points can cause serious problems in regression analysis. The statistics for identifying outlier and influential observations have been adequately studied in linear and linear mixed models and are available for users in most statistical packages. However, not much work has been done on similar methods for the analysis of multivariate survival data. In this chapter, we use martingale-based residuals to derive outlier and influence statistics for multivariate survival data model. We evaluate performance of the proposed statistics using simulation studies. Upon applying the proposed statistics to child survival data from Malawi, in which children were studied in 56 subdistricts, the outlier statistic detected five subdistricts as outliers to under-five mortality, while the influence statistic identified six subdistricts as having influence on the estimate of effect of being female on child survival, depending on the covariates used in the modelling process.

T. M. Kaombe (✉)
Department of Mathematical Sciences, School of Natural and Applied Sciences, University of Malawi, Zomba, Malawi
e-mail: tkaombe@unima.ac.mw

S. O. M. Manda
Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa
e-mail: Samuel.Manda@up.ac.za

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_15

1 Introduction

A statistical model becomes applicable when it fits the data well, and this is assessed through diagnostic statistics. They typically examine whether the fitted model complies with its assumptions and explore if there are observations that do not conform with the model. These assessments are usually done using the statistics such as residual, leverage, and others that quantify changes in estimates of coefficients upon dropping each data record, among other approaches (Yang, 2012). The statistics for identifying unusual observations to the fitted model are well-developed for linear and linear mixed models (Cook, 1977), and for some non-linear mixed models, such as binary logistic mixed model and Poisson mixed models (Sarkar et al., 2011; Zhang et al., 2016). However, there is deficiency of similar tools, when one intends to assess unusual observations upon fitting a multivariate survival model to data.

In survival data analysis, the outcome of interest is the time duration from some time origin until the occurrence of the event of interest. In medical and epidemiological research, the time period could be from initiation on antiretroviral treatment (ART) to loss to follow-up in a cohort of patients enrolled in a HIV pharmacological study; the number of months at death since birth in children under the age of five years; and the time to leukaemia relapse after bone marrow transplantation. In the case that the HIV pharmacological study has several sites, with each site having several HIV patients enrolled; or the case of several children spread across communities; and leukaemia cancer patients treated across several cancer clinics, then we have multivariate survival outcomes. Thus, multivariate survival data arise when subjects are grouped in some way, or when each individual experiences failure more than once, for example, leukaemia relapses over time (Ha et al., 2011; Manda, 2011; Maia et al., 2014). For the purpose of this chapter, we may also use clustered survival data for multivariate survival data. In contrast to univariate survival data, independence between survival times cannot be assumed for multivariate survival data. Ignoring the data correlations in the modelling, when they exist, may lead to under-estimation of variance and standard errors of regression coefficients and hence biased estimates (Liang & Zeger, 1993; Manda, 2011).

Multivariate survival models are therefore used to model the survival data that are clustered (Vaida & Xu, 2000). To fully examine the multivariate survival model, it becomes crucial to identify data points or groups of data points that might be outliers to the model or that might have overly influence on the model's inferences. This may help in improving the fit of the model to data and consequently give accurate predictions and conclusions from the model. This chapter therefore derives diagnostic statistics for detecting outlying and influential clusters of observations, upon fitting a multivariate survival model to the clustered data. The performance of the derived statistics is evaluated through simulations studies. The statistics are further illustrated through examples of their use on clustered child survival data from a health survey in Malawi. Malawi is located in south-eastern Africa. It borders

Tanzania to the north, Zambia to the west, and Mozambique to its east, south, and west. It has a population of just over 17 million people, and it is divided into 28 administrative districts (Malawi National Statistical Office (NSO), 2019). The rural and urban parts of these districts were studied as clusters, whose outlying patterns and influence on child survival was analysed.

1.1 The Multivariate Survival Model and Its Estimation

For censored survival outcomes t that are observed in M distinct clusters, the conditional Cox proportional hazard model, denoted by $h_{ij}(t_{ij}|\beta, b_i)$ (Abrahantes & Burzykowski, 2005; Xu et al., 2009), is given by:

$$h_{ij}(t_{ij}|\beta, b_i) = h_0(t)exp(X_{ij}^T\beta + Z_i^T b_i), \tag{1}$$

where $i = 1, 2, \ldots, M$ denotes clusters; $j = 1, 2, \ldots, n_i$ the study subjects in i-th cluster; and $h_0(t)$ is baseline hazard function. While X_{ij} is $p \times 1$ vector of fixed covariates; and β is $p \times 1$ vector of fixed regression coefficients. Further, Z_{ij} denotes $q \times 1$ vector of covariates that have random effects; and b_i is $q \times 1$ vector of random coefficients. In most cases, the random effects b_i's are assumed to be iid multivariate normal random variables, i.e. $b_i \sim N(\mathbf{0}, \mathbf{D})$, with $\mathbf{D}$ as $q \times q$ diagonal covariance matrix with similar entries at each level of b_i (Ripatti & Palmgren, 2000). So, we treat the random effects as normal random variables in this chapter. Furthermore, the censoring status is represented by δ_{ij}, which equals 1 for a subject that experienced the event of interest, and 0 otherwise.

The model (1) can estimate fixed effects β, predicted values of random effects b_i for each cluster, and amount of variation in survival times attributable to clustering of observations. This is usually achieved by computing marginal likelihood for β, upon treating b_i as nuisance parameters (Manda, 2001). Others use joint partial likelihood estimation, where estimates for both β and b_i are solved at the same time from a product of conditional likelihood function of data and the likelihood for random effects (Ripatti & Palmgren, 2000). We engaged the latter approach in this chapter. In addition, we apply the special case of model (1), called shared frailty model with fixed covariates, where random covariates Z_i are unity (Ripatti & Palmgren, 2000). We will also denote failure event-times in each cluster by $r_i = \sum_{j=1}^{n_i} d_{ij}$ and the risk set of individuals at time t_{ij} by $R(t_{il})$, $l = 1, 2, \ldots, r_i$ (Cox, 1972). Therefore, the conditional partial likelihood for observations in i-th cluster is given by:

$$L_i(\beta|b_i, t_{ij}, X_{ij}) = \prod_{l=1}^{r_i}[\frac{exp(X_{il}^T\beta + b_i)}{\sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)}]^{\delta_{ij}}. \tag{2}$$

Following the assumed normal distribution for random effects b_i, the complete joint partial likelihood function for the observed data and the random effects will take the form:

$$
\begin{aligned}
L(\beta, b_i) &= L_i(\beta|b_i, t_{ij}, X_{ij}) \times \prod_{i=1}^{M} f(b_i|\sigma^2) \\
&= \prod_{i=1}^{M}\prod_{j=1}^{n_i}\left[\frac{exp(X_{il}^T\beta + b_i)}{\sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)}\right]^{\delta_{ij}} \\
&\quad \times \prod_{i=1}^{M}\left[(2\pi)^{\frac{-1}{2}}\sigma^{2\,\frac{-1}{2}} exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{M} b_i^2)\right].
\end{aligned} \tag{3}
$$

This gives the joint partial log-likelihood function as:

$$
l(\beta, b_i) = \sum_{i=1}^{M}\sum_{j=1}^{n_i}[X_{ij}^T\beta + b_i - ln \sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)] + ln[(2\pi)^{\frac{-M}{2}}(\sigma^2)^{\frac{-M}{2}}] - \frac{1}{M/(2\sigma^2)}\sum_{i=1}^{M} b_i^2. \tag{4}
$$

From the log-likelihood (4), we obtain the vectors of score functions for the fixed parameters β and for random parameters b_i as follows:

$$
U_\beta = \frac{\partial l(\beta, b_i)}{\partial \beta} = \sum_{j=1}^{n_i}\sum_{i=1}^{M}\delta_{ij}\left[X_{ij} - \frac{\sum_{s\in R(t_{il})} X_{is} exp(X_{is}^T\beta + b_i)}{\sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)}\right], \tag{5}
$$

and

$$
U_{b_i} = \frac{\partial l(\beta, b_i)}{\partial b_i} = \sum_{j=1}^{n_i}\sum_{i=1}^{M}[1 - \frac{\sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)}{\sum_{s\in R(t_{il})} exp(X_{is}^T\beta + b_i)}] - \frac{M}{\sigma^2}\sum_{i=1}^{M} b_i. \tag{6}
$$

The partial derivatives in equations (5) and (6) will yield $(p + 1) + q$ equations in $(p + 1) + q$ unknowns β and b_i. Hence, unique maximum likelihood estimators for the parameters can be obtained by solving for these parameters when the equations (5) and (6) are equated to zero. Due to intractable nature of the equations (5) and (6), numerical techniques are used to obtain the solutions (Ripatti & Palmgren, 2000).

In the next section, we present a derivation of outlier detection statistic in the analysis of multivariate survival data model (1), with a simulation study and application example. In Sect. 3, we derive an influence statistic for the same multivariate survival data model. We also include a simulation study and an application example. Conclusions are given in Sect. 4.

2 Outlier Analysis for Multivariate Survival Data

As stated in Sect. 1, the subject of model outliers has been widely studied in linear mixed-effects models. The outlier measures help to detect data records or groups

of data records that deviate much from the fitted model. Such measures typically include a residual $\widehat{\mathbf{e}}$, given by:

$$\widehat{\mathbf{e}} = \mathbf{y} - (\mathbf{X}\widehat{\beta} + \mathbf{Z}\widehat{b}), \tag{7}$$

where $\mathbf{y}$ is the response vector in linear mixed model, $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}b + \epsilon$, with $\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 I)$, $b \sim N(\mathbf{0}, \mathbf{D})$; and $\widehat{\beta}$ and $\widehat{b}$ are estimates of the model's respective fixed and random effects. The probability distribution of the residual (3) is usually normal with mean zero and constant variance. Hence, the values in the periphery of the distribution of the residual are considered outliers.

However, the variance of the residual (8) is biased estimator of variance of the unknown error term, which causes its distribution to be skewed and not adequate for outlier analysis on both sides of the mean of the residual (8). For this reason, the standardised residual is used for outlier analysis as it has a symmetric shape about mean zero. This is given by:

$$\lambda = \widehat{\mathbf{e}}/stdev(\widehat{\mathbf{e}}), \tag{8}$$

where $stdev(\widehat{\mathbf{e}})$ is standard deviation of $\widehat{\mathbf{e}}$. Both residuals (7) and (8) detect single observations outliers to a linear mixed-effects model. Langford and Lewis (1998) extended the use of (8) to examining outlying groups of data to linear mixed model. When plotted against each cluster, the clusters in which the plots of the standardised residual (8) are much polarised from others are regarded as outliers to the linear mixed-effects model (Langford & Lewis, 1998).

The main weakness of method of assessing group outliers by plotting single observations standardised residuals against clusters is that when the plots of standardised residuals for various clusters are highly overlapping, the method is no longer reliable in examining group outliers (Kaombe & Manda, 2022). In the next subsection, we present an outlier detection statistic for multivariate survival model that applies to natural groups of data points.

2.1 Proposed Outlier Statistic for Multivariate Survival Data

Through extending the work of Therneau et al. (1990) on residual for univariate survival data (also called martingale residual), a residual for single observations in the survival mixed model (1) is defined in Kaombe and Manda (2022) as:

$$m(t_{il}) = N(t_{il}) - \int_0^{t_{il}} Y_{il}(t) exp(X_{il}^T \widehat{\beta} + \widehat{b}_i) d\widehat{H}_0(t), \tag{9}$$

where $N(t_{il})$ is number of events observed over time t_{il}; $Y_{il}(t)$ is an indicator variable showing if ij-th unit is at risk at time t_{ij}; and $\widehat{H}_0(t)$ is cumulative baseline

hazard function. The residual (10) measures the difference between observed and estimated number of events at specific time points over $[0, t]$ (Therneau et al., 1990). For model (1), which is time-independent covariates model, this residual becomes (Kaombe & Manda, 2022):

$$m(t_{il}) = N(t_{il}) - \widehat{H}_0(t)exp(X_{il}^T\widehat{\beta} + \widehat{b}_i)$$

$$\Rightarrow \begin{bmatrix} m(t_{11}) \\ \vdots \\ m(t_{1n_1}) \\ m(t_{21}) \\ \vdots \\ m(t_{2n_2}) \\ \vdots \\ m(t_{M1}) \\ \vdots \\ m(t_{Mn_M}) \end{bmatrix} = \begin{bmatrix} N(t_{11}) - \widehat{H}_0(t)exp(X_{11}^T\widehat{\beta} + \widehat{b}_1) \\ \vdots \\ N(t_{1n_1}) - \widehat{H}_0(t)exp(X_{1n_1}^T\widehat{\beta} + \widehat{b}_1) \\ N(t_{21}) - \widehat{H}_0(t)exp(X_{21}^T\widehat{\beta} + \widehat{b}_2) \\ \vdots \\ N(t_{2n_2}) - \widehat{H}_0(t)exp(X_{2n_2}^T\widehat{\beta} + \widehat{b}_2) \\ \vdots \\ N(t_{M1}) - \widehat{H}_0(t)exp(X_{M1}^T\widehat{\beta} + \widehat{b}_M) \\ \vdots \\ N(t_{Mn_M}) - \widehat{H}_0(t)exp(X_{Mn_M}^T\widehat{\beta} + \widehat{b}_M) \end{bmatrix}. \tag{10}$$

The distribution of residual (11) is negatively skewed in each cluster, since $N(t_{il}) \in [0, 1]$ and $\widehat{H}_0(t)exp(X_{il}^T\widehat{\beta} + \widehat{b}_i)$ has values in $[0, \infty)$. Hence, its use to directly assess outlier single observations is only applicable for observations that failed too late, and not the ones that failed too early (Therneau et al., 1990).

For this reason, its standardised form called deviance residual is used (Therneau et al., 1990). The deviance residual measures disagreement between element of log-likelihood of the fitted model and corresponding point of the log-likelihood that would result if each observation were fitted exactly (Therneau et al., 1990; Sarkar et al., 2011). The extension of deviance residual for model (1) is given in Kaombe and Manda (2022) by:

$$d_{il} = sign(m(t_{il}))[-2(m(t_{il}) + \delta_{il}log(\delta_{il} - m(t_{il})))]^{\frac{1}{2}}$$

$$\Rightarrow \begin{bmatrix} d_{11} \\ \vdots \\ d_{1n_1} \\ d_{21} \\ \vdots \\ d_{2n_2} \\ \vdots \\ d_{M1} \\ \vdots \\ d_{Mn_M} \end{bmatrix} = \begin{bmatrix} sign(m(t_{11}))[-2(m(t_{11}) + \delta_{11}log(\delta_{11} - m(t_{11})))]^{1/2} \\ \vdots \\ sign(m(t_{1n_1}))[-2(m(t_{1n_1}) + \delta_{1n_1}log(\delta_{1n_1} - m(t_{1n_1})))]^{1/2} \\ sign(m(t_{21}))[-2(m(t_{21}) + \delta_{21}log(\delta_{21} - m(t_{21})))]^{1/2} \\ \vdots \\ sign(m(t_{2n_2}))[-2(m(t_{2n_2}) + \delta_{2n_2}log(\delta_{2n_2} - m(t_{2n_2})))]^{1/2} \\ \vdots \\ sign(m(t_{M1}))[-2(m(t_{M1}) + \delta_{M1}log(\delta_{M1} - m(t_{M1})))]^{1/2} \\ \vdots \\ sign(m(t_{Mn_M}))[-2(m(t_{Mn_M}) + \delta_{Mn_M}log(\delta_{Mn_M} - m(t_{Mn_M})))]^{1/2} \end{bmatrix}. \tag{11}$$

The standardisation helps to symmetrise values of the extended martingale residual (11) because the logarithm applied on (11) inflates values of (11) that are near 1, while the square root contracts large negative values (Therneau et al., 1990). Thus, the approach of Langford and Lewis (1998) on assessing group residuals graphically through observing skewness of standardised residual can be applied on the deviance residual (12).

Owing to the weaknesses of the approach suggested in Langford and Lewis (1998), the work of Kaombe and Manda (2022) proposed a group outlier statistic based a residual that estimates poorly fitted cluster as a whole to the multivariate survival model (1). The statistic, denoted by k, is a ratio of within-cluster variance of extended deviance residual (12) to between-cluster variance (Kaombe & Manda, 2022), given by:

$$\begin{aligned} k &= \frac{1}{L}(k_1, \ldots, k_M)^T \\ &= \frac{1}{L}\left(\frac{\sum_{j=1}^{n_1}(d_{1j} - \bar{d}_1)^2}{n_1 - 1}, \ldots, \frac{\sum_{j=1}^{n_M}(d_{Mj} - \bar{d}_M)^2}{n_M - 1}\right)^T. \end{aligned} \tag{12}$$

where $L = \frac{\sum_{i=1}^{M} n_i(\bar{d}_i - \bar{\bar{d}})^2}{M-1}$ is between-cluster variance of d_{ij}; $\bar{\bar{d}} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i} d_{ij}}{n}$ is the grand mean of the deviance residual d_{ij}; and $\bar{d}_i = \frac{\sum_{j=1}^{n_i} d_{ij}}{n_i}$ is the mean of d_{ij} for any fixed i; $n = n_1 + n_2 + \ldots + n_M$ is number of subjects in entire dataset. The outlier statistic (13) is an $M \times 1$ vector that measures overall sparseness of survival times of the subjects in each cluster off the fitted survival curve (Kaombe & Manda, 2022). As the fitted survival curve will have to pass through all clusters, small values of the statistic k in (13) will go with clusters that are well-fitted by the model and have observations that closely span the fitted survival curve. On the other hand, large values of the statistic (13) will show outlying clusters with poorly fitted observations (Kaombe & Manda, 2022).

Some of the properties of the statistic k are that $k \in [0, \infty)$ and it is a non-linear function, because it is computed from variances that have support $[0, \infty)$ (Kaombe & Manda, 2022). The expected value of k can be computed using second order Taylor series expansion about $\mu = (\mu_{k_i}, \mu_l)$ (Van Kempen & Van Vliet, 2000) as:

$$\begin{aligned} E(k) &= E(K_i/L) \\ &= E(f(K_i, L)) \\ &\approx E[f(\mu) + f'_{k_i}(\mu)(k_i - \mu_{k_i}) + f'_l(\mu)(l - \mu_l) \\ &\quad + \frac{1}{2}\{f''_{k_i k_i}(\mu)(k_i - \mu_{k_i})^2 + 2f''_{lk_i}(\mu)(k_i - \mu_{k_i})(l - \mu_l) + f''_{ll}(\mu)(l - \mu_l)^2\}] \\ &= f(\mu) + \frac{1}{2}\{f''_{k_i k_i}(\mu)Var(K_i) + 2f''_{lk_i}(\mu)Cov(L, K_i) + f''_{ll}(\mu)Var(L)\} \\ &= \frac{\mu_{k_i}}{\mu_l} - \frac{1}{\mu_l^2}Cov(K_i, L) + \frac{\mu_{k_i}}{\mu_l^3}Var(L), \end{aligned} \tag{13}$$

where $f(\mu) = \mu_{k_i}/\mu_l$, $f''_{k_i k_i}(\mu) = 0$, $f''_{lk_i}(\mu) = -1/(\mu_l)^2$, and $f''_{ll}(\mu) = 2\mu_{k_i}/(\mu_l)^3$, since $f(K_i, L) = K_i/L$ and $E(K_i/L) = E(f(K_i, L))$. Also, $E(k_i - \mu_{k_i}) = E(l - \mu_l) = 0$; $Var(K_i) = E(k_i - \mu_{k_i})^2$, and $Cov(K_i, L) = E[(k_i - \mu_{k_i})(l - \mu_l)]$. Whereas, variance of k can be estimated using first-order Taylor series expansion of $f(K_i, L)$ around $\mu = (\mu_{K_i}, \mu_l)$ as

$$\begin{aligned} Var(k) &= Var(K_i/L) \\ &= Var(f(K_i, L)) \\ &= E\{[f(K_i, L) - f(\mu)]^2\} \\ &\approx E\{[f(\mu) + f'_{k_i}(\mu)(k_i - \mu_{k_i}) + f'_l(\mu)(l - \mu_l) - f(\mu)]^2\} \\ &= f'^2_{k_i}(\mu)Var(K_i) + 2f'_{k_i}(\mu)f'_l(\mu)Cov(K_i, L) + f'^2_l(\mu)Var(L) \\ &= \frac{1}{\mu_l^2}Var(K_i) - 2\frac{\mu_{k_i}}{\mu_l^3}Cov(K_i, L) + \frac{\mu_{k_i}^2}{\mu_l^4}Var(L). \end{aligned} \tag{14}$$

The outlier statistic k in Eq. (13) can be analysed using graphical methods as is done in linear mixed models. Usually, the value of a residual for an observation or group of observations is assessed in comparison with the others to detect outlying data or groups of data (Zewotir & Galpin, 2007; Kaombe & Manda, 2022).

This section has presented an outlier statistic for examining outlying clusters of observations to a fitted multivariate survival data model. In the next section, we present a simulation study to exemplify its use.

2.2 Simulation Study

We simulated survival times data from $T \sim Exp(1)$ using cumulative hazard inversion method (Brilleman et al., 2018) to evaluate performance of the outlier statistic (13). The model used is:

$$h_{ij}(t|b_i, X_{ij}) = h_0(t)exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + b_i), \tag{15}$$

where $h_0(t) = 0.1$; $X_1 \sim Bernoulli(0.7)$; $\beta_1 = 0.5$; $X_2 \sim N(0, 1)$; $\beta_2 = 1$; and $b_i \sim N(0, 0.4^2)$. The inversion method obtains t_{ij} from $H_{ij}^{-1}(-log(S(t_{ij})))$, having $S(t_{ij}) \sim Uniform(0, 1)$, such that $H_{ij}(t) = -log(Uniform(0, 1)) \sim Exp(1)$. The censoring statuses were drawn from $Ber(0.4)$. The generation was done with the help of `R package simsurv` (Moriña & Navarro, 2014; Brilleman et al., 2018). The sets of samples had 10, 20, and 50 clusters, each with 80 and 500 subjects per cluster. This assessed impact of sample size per cluster as well as number of clusters per dataset on the performance of the proposed outlier statistic (13). There

were 100 and 1000 replications involved per each case of generated data (Kaombe & Manda, 2022).

The evaluation involved perturbing data in first two clusters and observing if the statistic (13) could detect it upon fitting model (16) to the data (Kaombe & Manda, 2022). The perturbations included: $b_{1,2} \sim N(10, 2.5^2)$, $N(15, 5.5^2)$, leaving β_1 and β_2 the same; and $\beta_1 = 1.8, 2.7$ followed by $\beta_2 = 2.0, 2.5$, without adjusting the other parameters. Finally, both $\beta_1 = 1.8, 2.7$ and $\beta_2 = 2.0, 2.5$, with b_i intact, then, $\beta_1 = 1.8, 2.7$, $\beta_2 = 2.0, 2.5$ and $b_{1,2} \sim N(10, 2.5^2)$, $N(15, 5.5^2)$ (Kaombe & Manda, 2022). This was done because the outlying tendency of a cluster can be as a result of a combination of values for fixed covariates, survival outcomes, and random effects in the model (Zewotir & Galpin, 2006). We assessed the proportion per total 100 or 1000 simulations in which the proposed outlier statistic (13) correctly detected cluster 1 and 2 as unusual at cutoffs of mean [maximum $(k_i : i = 3, 4, \ldots, M)$] and mean [minimum $(k_i : i = 3, 4, \ldots, M)$] (Xiang et al., 2002; Kaombe & Manda, 2021). The higher the percentage of correct identification of problematic cluster 1 or 2 by the proposed outlier statistic (13), the more effective the statistic (Kaombe & Manda, 2022). We fitted model (16) to all cases of simulated data and analysed performance of the proposed outlier statistic (13) as per the above procedure.

2.2.1 Simulation Results

Figure 1 shows outlier statistic results for data from two simulation scenarios, in which the first involved perturbed random effects, 1 (a) and the second perturbed fixed effects 1 (b). It was shown that the proposed outlier statistic had detected

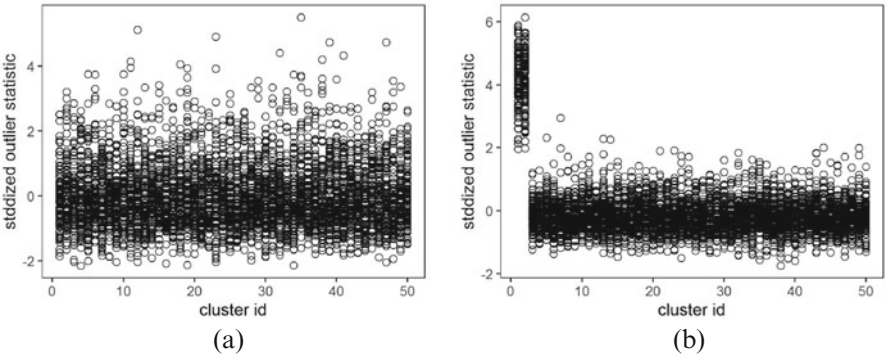


Fig. 1 Plots of the proposed outlier statistic for two simulation cases when perturbed models were used in first two clusters. (**a**) Plots of outlier statistic for a case of data with perturbed $b \sim (15, 5.5^2)$ in 2 of 50-clusters sample, each with 80 subjects and with 100 replications. (**b**) Plots of outlier statistic for a case of data with perturbed $\beta_1 = 2.7$ in 2 of 50-clusters sample, each with 500 subjects and with 100 replications. Source: Researcher

perturbations done to clusters 1 and 2 for the data involving perturbed fixed effects parameters, but not perturbed random effects. The graph showed that values of the outlier statistic were exceedingly higher in clusters 1 and 2 than the rest clusters in Fig. 1b, where β_1 was perturbed in the first two clusters, and the values did not show a different pattern for all clusters in Fig. 1a that had perturbed random effects.

The results on overall performance of the proposed outlier statistic are given in Table 1 for cases of separate perturbations and in Table 2 for the data that involved joint perturbed parameters in first two clusters. In each case, the results showed that the statistic was effective in detecting the outlying clusters in situations that involved perturbed fixed effects, but not random effects. From Table 1, it is shown that the statistic correctly detected the first two clusters with high rates of up to 100% of the times, when β_1 or β_2 was perturbed. When it is the random effects only that were adjusted in cluster 1 and 2, the success rates of the statistic were all zero. This means that the values of the outlier statistic for cluster 1 and 2 were not different from the rest cluster if it is the random effects that were perturbed in the first two clusters. This is due to the fact that observations in the same cluster share a random effect, which might contribute less to within-cluster variation of the deviance residual that is used in the proposed outlier statistic, unlike fixed covariates values that vary from subject to subject even within the same cluster (Kaombe & Manda, 2022).

The results in Table 1 also showed that the outlier statistic performed better when sample size per cluster increased from 80 to 500, and when the value of perturbed fixed effect increased. Further, the performance of the statistic was the same between 100 and 1000 simulations where large cluster sample sizes were used. But, the success rates dropped slightly at 1000 simulation in situations that involved small sample sizes per cluster. Finally, the outlier statistic performed equally across different numbers of clusters per data set, holding constant the cluster sample size and fixed effect size.

As for the cases of data from jointly perturbed parameters given in Table 2, it is shown that the outlier statistic correctly detected the outlying clusters 1 and 2 a minimum of 39.1% and up to 100% of the times, when β_1 and β_2 were jointly perturbed. The detection rates ranged from 38.3% and up to 100% of the times, when β_1, β_2 and b_j were jointly adjusted. This means that performance of the statistic was insensitive of involvement of random effects in the joint perturbations. This agrees with results in Table 1 where values of the outlier statistic in clusters 1 and 2 were consistent with those in the rest clusters, although the random effects were perturbed in the first two clusters.

Like in the case of separate perturbations, the performance of the statistic was not different between 100 and 1000 simulations in cases where cluster size was large. Similarly, there was improved performance of the statistic when the sample size per cluster increased from 80 to 500, and when the value of perturbed fixed effect increased.

Table 1 Percentage of times per 100 or 1000 simulations[a] in which cluster 1 or 2 was detected as outlier by the proposed statistic; a case of separate perturbations to β_1, β_2, or $b_{1,2}$ under 10, 20, or 50 clusters per dataset, each with 80 or 500 subjects

					100 replicates		1000 replicates	
M	n_i	β_1	β_2	$b_{1,2}$	%Cluster1	%Cluster2	%Cluster1	%Cluster2
10	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
10	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
20	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
20	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
50	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
50	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
10	80	1.8	1	$N(0, 0.4^2)$	0	0	0	0
	80	2.7	1	$N(0, 0.4^2)$	20	17	22	22
10	500	1.8	1	$N(0, 0.4^2)$	50	50	25.8	24.9
	500	2.7	1	$N(0, 0.4^2)$	100	100	88.9	89.6
20	80	1.8	1	$N(0, 0.4^2)$	3	3	0.6	0.4
	80	2.7	1	$N(0, 0.4^2)$	17	7	2.2	1.7
20	500	1.8	1	$N(0, 0.4^2)$	84	83	17.4	18.9
	500	2.7	1	$N(0, 0.4^2)$	96	100	95.2	95.5
50	80	1.8	1	$N(0, 0.4^2)$	11	15	8.0	8.2
	80	2.7	1	$N(0, 0.4^2)$	6	2	0.7	0.7
50	500	1.8	1	$N(0, 0.4^2)$	57	59	31.7	34.6
	500	2.7	1	$N(0, 0.4^2)$	100	98	98.5	97.5
10	80	0.5	2.0	$N(0, 0.4^2)$	95	92	57	59.6
	80	0.5	2.5	$N(0, 0.4^2)$	100	100	92.7	93.8
10	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.9	99.6
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	100	100
20	80	0.5	2.0	$N(0, 0.4^2)$	82	83	72.8	74.7
	80	0.5	2.5	$N(0, 0.4^2)$	99	98	92.5	93
20	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.7	100
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	99.7	99.7
50	80	0.5	2.0	$N(0, 0.4^2)$	93	89	79.6	80.4
	80	0.5	2.5	$N(0, 0.4^2)$	98	97	87.2	87.8
50	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.1	99.1
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	99.9	99.8

[a] No perturbations were done to data in other clusters than 1 and 2, in those other clusters model (16) had $\beta_1 = 0.5$, $\beta_2 = 1$, and $b_i \sim N(0, 0.4^2)$

Table 2 Percentage of times per 100 or 1000 simulations[a] in which cluster 1 or 2 was detected as outlier by the proposed statistic; a case of joint perturbations among β_1, β_2, and $b_{1,2}$ under 10, 20 or 50 clusters per dataset, each with 80 or 500 subjects

					100 replicates		1000 replicates	
M	n_i	β_1	β_2	$b_{1,2}$	%Cluster1	%Cluster2	%Cluster1	%Cluster2
10	80	1.8	2.0	$N(0, 0.4^2)$	55	68	39.1	41.8
	80	2.7	2.5	$N(0, 0.4^2)$	90	89	74.6	72.7
10	500	1.8	2.0	$N(0, 0.4^2)$	100	100	99.4	99.1
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	100	100
20	80	1.8	2.0	$N(0, 0.4^2)$	58	58	43.4	44.8
	80	2.7	2.5	$N(0, 0.4^2)$	86	88	70.2	69.6
20	500	1.8	2.0	$N(0, 0.4^2)$	99	99	98.9	98.6
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	100	100
50	80	1.8	2.0	$N(0, 0.4^2)$	59	54	41.2	39.7
	80	2.7	2.5	$N(0, 0.4^2)$	86	87	69	66.8
50	500	1.8	2.0	$N(0, 0.4^2)$	100	100	99	98.5
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	97.7	98
10	80	1.8	2.0	$N(10, 2.5^2)$	75	74	38.7	40.9
	80	2.7	2.5	$N(15, 5.5^2)$	85	86	74.6	72.6
10	500	1.8	2.0	$N(10, 2.5^2)$	100	100	99.3	99.2
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	100	100
20	80	1.8	2.0	$N(10, 2.5^2)$	48	54	38.3	35.4
	80	2.7	2.5	$N(15, 5.5^2)$	82	76	73.4	73.3
20	500	1.8	2.0	$N(10, 2.5^2)$	100	100	97.9	98.4
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	99.5	99.3
50	80	1.8	2.0	$N(10, 2.5^2)$	65	51	41.6	40.8
	80	2.7	2.5	$N(15, 5.5^2)$	79	80	68.9	71.8
50	500	1.8	2.0	$N(10, 2.5^2)$	100	99	96.1	95.2
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	98.7	99.4

[a] No perturbations were done to data in other clusters than 1 and 2, in those other clusters model (16) had $\beta_1 = 0.5$, $\beta_2 = 1$, and $b_i \sim N(0, 0.4^2)$

2.3 Application to Malawi Child Survival Data

This section presents the implementation of the proposed outlier statistic demonstrated in Kaombe and Manda (2022), which used under-five mortality data from 2015–16 Malawi Demographic and Health Survey (MDHS) data. The survey took place between 19 October 2015 and 18 February 2016 and had a sample of 17,286 children, whose information was provided by women respondents and caregivers aged 15 to 49 years. It used two-stage stratified sampling, with enumeration areas (EAs) as primary and households as secondary sampling units (Malawi National Statistical Office (NSO) and ICF, 2017). The data can be accessed at

Table 3 Data summary for child survival study using 2015–16 MDHS

Variable	Level	Number of births	%Died
Overall sample	All	14,645	3.68
Sex of child	Male	7,393	4.03
	Female	7,252	3.32
Birth order of child	1	3,769	4.48
	2–3	5,480	3.19
	≥ 4	5,396	3.61
Place of birth	Home or other	369	4.07
	Hospital	14,276	3.67
Place of residence	Urban	2,544	3.03
	Rural	12,101	3.82
Birth weight (kgs)	=>2,500g	12,465	3.45
	<2,500g	2,180	5.00
Household wealth index	Poor	6,166	3.97
	Middle	2,823	3.54
	Rich	5,656	3.43
Mother's education	No formal education	1,588	3.46
	Primary	9,611	3.85
	Secondary or above	3,446	3.31

www.DHSprogram.com. There were 56 clusters that were studied, which represented each district's rural and urban subdistricts.

The analysis involved a survival frailty model using death of a child from any cause before age of 60 months as event of interest, and age at death or at censoring time as event-time. Children whose event-times were zero months were assigned random values between 0 and 1 from a uniform probability distribution to reflect proportion of month-days for their event-times. The children who were still alive during survey time or had survived up to 60 months were censored. Due to missing data in some key variables that were used in the model, such as birth weight, the analysis of outlier clusters to child mortality involved complete cases of 14,645 children. The data and the rest of the variables that were analysed are given in Table 3.

2.3.1 Results on Effects of Some Variables on Under-Five Child Mortality

The model results in Table 4 showed that female sex, birth order of 2 and above had significantly lower logarithm of hazard of mortality compared to their reference groups. While, the birth weight of below 2,500 grams was associated with increased logarithm of hazard of death in under-five children compared to birth weight of 2,500 grams and above. Further, the results showed that place of residence, household wealth, place of birth, and mother's education had no significant effect

Table 4 Estimates from the fitted frailty hazard model on child survival data

Variable	Level	Coefficient (p-value)	Coefficient (p-value)
		Model 1	Model 2
Sex of child	Male (ref)		
	Female	−0.235 (0.0069)	−0.234 (0.0070)
Birth order of child	1		
	2 − 3	−0.335 (0.0020)	−0.328 (0.0024)
	≥4	−0.238 (0.0307)	−0.217 (0.0394)
Birth weight (kgs)	=>2,500g (ref)		
	<2,500g	0.602 (<0.0001)	0.615 (<0.0001)
Mother's education	No formal education		
	Primary	0.110 (0.4572)	
	Secondary or above	0.020 (0.9141)	
Place of birth	Home or other (ref)		
	Hospital	−0.023 (0.9298)	
Place of residence	Urban (ref)		
	Rural	0.201 (0.1633)	
Household wealth index	Poor		
	Middle	−0.097 (0.4194)	
	Rich	−0.035 (0.7537)	
Variance of random effects	Cluster	0.0228 (0.1875)	0.0257 (0.1617)

on log-hazard of death of a child. These were dropped in the ultimate models that were used in this study. It was also shown that the variation between clusters was significantly different from 0 for reduced models with significant covariates only. This implied that time to death of a child differed significantly between clusters.

Using the reduced model, with three significant covariates, the values of the proposed outlier statistic were computed to analyse outlying subdistricts to under-five mortality. This was done along with the visual inspection approach for standardised residuals suggested in Langford and Lewis (1998). The baseline hazard of 63 deaths per 1000 live births, which was the prevailing national under-five mortality rate, was used (Malawi National Statistical Office (NSO) and ICF, 2017).

2.3.2 Results for Outlier Subdistricts on Under-Five Mortality

The results in Fig. 2a indicate that the proposed outlier statistic had detected five subdistricts as outliers to child mortality, these are: *Likoma* rural, *Lilongwe* urban, *Chikwawa* rural, *Dedza* rural, and *Neno* rural. This means that, these subdistricts were poorly fitted by the child survival model, and they had a different pattern of mortality compared to the rest subdistricts. Using the method of visual inspection from linear mixed model, as in Fig. 2b we could not clearly identify an outlying subdistrict due to overlaps of the plots.

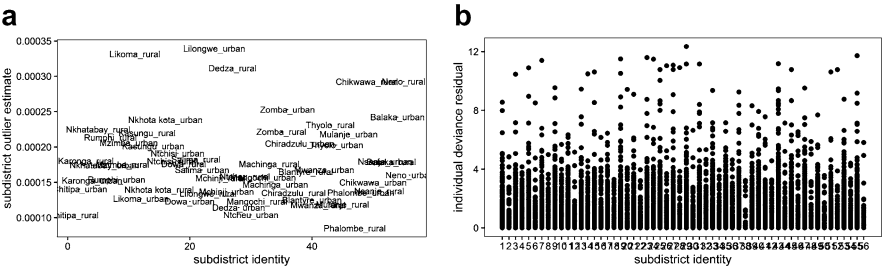


Fig. 2 Outlier assessment results using the proposed group outlier statistic in comparison with method of visual inspection of standardised residuals (Langford & Lewis, 1998) applied on Malawi child survival data, 2015–16 MDHS. (**a**) Estimates of proposed outlier statistic per subdistrict upon fitting the frailty Cox model on child survival data. (**b**) Plots of deviance residuals for children in each cluster following a frailty model on child survival data. Source: Researcher

3 Influence Analysis for Multivariate Survival Data

We present a derivation of an influence statistic for clustered data upon fitting a multivariate survival model, based on the work of Kaombe and Manda (2021). This is preceded by a review of influence statistics for univariate survival model.

3.1 Some Common Influence Statistics for Univariate Survival Data

Given maximum likelihood (ML) estimators $\widehat{\theta}$ for model parameters $\theta = \{\beta, b_i, D\}$ and $\widehat{\theta}_{(ij)}$ as the estimator obtained upon deleting ij-th observation from data. The influence of ij-th data point on $\widehat{\theta}$ is defined as the difference in estimators, $\Delta\widehat{\theta}_{ij} = \widehat{\theta} - \widehat{\theta}_{(ij)}$ (Das & Gogoi, 2015; Cain & Lange, 1984). The estimates for $\Delta\widehat{\theta}_{ij}$ can be manually computed upon deleting each data record and refitting the model and calculate the difference in estimates from full and reduced data or get difference of one-step iterative approximations for non-linear models. However, the manual computations are time-consuming as they require refitting the model (n + 1) times to the dataset. This is the reason why efficient model post-estimation influence statistics, computed from one fitting of the model, are used.

A number of such techniques exist for linear models and are extended to univariate survival models. With generalised linear and linear mixed-effects models (glm and glmm), where parameter estimators $\widehat{\theta}$ are obtained analytically, influence statistic $\Delta\widehat{\theta}_{ij}$ is a function of model's basic building blocks, i.e. Studentized residuals, error contrast matrix, and inverse of covariance matrix of response variable (Zewotir & Galpin, 2005). In glm and glmm, $\Delta\widehat{\theta}_{ij}$ is either computed from methods like Cook's distance (Cook, 1977) or it is approximated for one-step ML estimation using updating formulae techniques (Zewotir, 2008; Nobre & Singer,

2011). Others use first-order Taylor series expansion on score functions about the estimator $\widehat{\theta}_{(ij)}$ (Xiang et al., 2002). However, with Cox proportional hazard (PH) model, the analytic case-deletion influence techniques, such as Cook's distance, do not apply, since subjects enter the likelihood as members of various risk sets, such that deleting a data record affects several risk sets rather than one (Cox, 1972). For this reason, various approximations for influence statistics have been developed for univariate survival data, which we present in this section.

One influence technique that is used for univariate survival data is the first-order Taylor series expansion of score function about a unity weight ϖ_{ij} for an observation. The weight $\varpi_{ij} = 0$ for a subject that has been removed from data and $\varpi_{ij} = 1$ otherwise (Cain & Lange, 1984). The subjects' weights ϖ_{ij} result into weighted partial likelihood $L(\beta(\varpi_{ij}))$, and weighted score function $U_{\beta(\varpi_{ij})}$ from the model. Subsequently, the weighted ML estimators $\widehat{\beta}(\varpi_{ij})$ become $\widehat{\beta}(1) = \widehat{\beta}$ or $\widehat{\beta}(0) = \widehat{\beta}_{(ij)}$, where $\widehat{\beta}_{(ij)}$ is the estimator obtained upon dropping ij-th case in the dataset, and $\widehat{\beta}$ the one obtained from full data. Then, the influence estimate is given by $\Delta\widehat{\beta}_{ij} = \widehat{\beta} - \widehat{\beta}_{(ij)} = \partial\widehat{\beta}/\partial\varpi_{ij}$, which is obtained by solving for $\partial\widehat{\beta}/\partial\varpi_{ij}$ when the score function is equated to zero (Cain & Lange, 1984), as follows:

$$\begin{aligned}(\partial U/\partial\widehat{\beta})(\partial\widehat{\beta}/\partial\varpi_{ij}) + \partial U/\partial\varpi_{ij} &= 0\\ \therefore \partial\widehat{\beta}/\partial\varpi_{ij} &= (-\partial U/\partial\widehat{\beta})^{-1}\partial U/\partial\varpi_{ij},\end{aligned} \tag{16}$$

where the likelihood for univariate model is: $L(\beta|\mathbf{t},\mathbf{X}) = \prod_r [\frac{exp(X_{ij}^T\beta)}{\sum_{s\in R(t_{il})}\varpi_{ij}exp(X_{is}^T\beta)}]^{\varpi_{ij}}$, and the weighted score function is first derivative of logarithm of $L(\beta|\mathbf{t},\mathbf{X})$ with respect to β. The method (17) is also referred to as infinitesimal jackknife statistic for influence of an observation (Therneau et al., 1990).

Another method is the score residual, which is essentially a product of a subject's residual and its extremity in covariate value (Therneau et al., 1990) given by:

$$v_{ij}(\widehat{\beta}) = \int_0^{\infty}[X_{ijp}(t) - \bar{X}_p(\widehat{\beta},t)]dm(t_{ij}), \tag{17}$$

where $m(t_{ij}) = N(t_{ij}) - \int_0^{t_{ij}} Y_{ij}(t)exp(X_{ij}^T(t)\widehat{\beta})d\widehat{H}_0(t)$ is residual of ij-th observation at time t_{ij}, also called martingale residual, which measures excess number of events; and p is number of covariates; while $\bar{X}_p = \frac{\sum X_{ijp}exp(X_{ij}^T\widehat{\beta})}{\sum_{s\in R(t_{il})}exp(X_{is}^T\widehat{\beta})}$ is the weighted average of covariate X_{ijp} over $R(t_{il})$ risk sets. The measure (18) is used to estimate sensitivity of log-likelihood to infinitesimal displacements of $\widehat{\beta}$. Using a weighted partial likelihood, Therneau et al. (1990) showed that the residual (18) is similar to the jackknife measure (17) and that $\partial U/\partial\varpi_{ij} = (v_{ij1}, v_{ij2}, \ldots, v_{ijp})^T$.

The third method is the augmented or perturbed regression model (Storer & Crowley, 1985; Therneau et al., 1990), which is a one-step update in $\widehat{\theta}$ when a single indicator covariate is added to the model. The added covariate has value 1 for ij-th data point and 0 for all other observations (Therneau et al., 1990). The augmented model influence statistic for univariate survival model (Storer & Crowley, 1985) is given by:

$$\begin{aligned} \widehat{\beta}_1 &= \widehat{\beta}_0 + I^{-1}(\widehat{\beta}_0)\dot{l}(\widehat{\beta}_0) \\ \Rightarrow \widehat{\beta}_1 - \widehat{\beta}_0 &= I^{-1}(\widehat{\beta}_0)\dot{l}(\widehat{\beta}_0) \\ &= \frac{-I^{-1}(\widehat{\beta}_0)\xi_{ij}}{\pi_{ij} - \xi_{ij}^T I^{-1}(\widehat{\beta}_0)\xi_{ij}} m(t_{ij}), \end{aligned} \tag{18}$$

where $m(t_{ij})$ is the martingale residual defined along with Eq. (8), $\xi_{ijp} = \widehat{H}_0(X_{ijp} - \bar{X}_p(\widehat{\beta}))exp(X_{ij}^T\widehat{\beta})$ represents a column vector from design matrix $\mathbf{X}$ corresponding to $1's$, $\pi_{ij} = \widehat{H}_0(t)(1 - \bar{c}_{ij}(\widehat{\beta}))exp(\widehat{\beta}^T X_{ij}^T)$ is the diagonal identity matrix with entries 1 throughout, except for the subject that has been removed, which has 0 entry, and c_{ij} is the indicator covariate that has been added to the dataset (Storer & Crowley, 1985).

These methods are all related, since they are functions of subject's leverage and residual measures. Moreover, Therneau et al. (1990) demonstrated that the three methods yield similar estimates of influence, but the score residual (18) has a number of advantages, including simplicity of interpretation. Hence, we applied the score residual technique to derive the influence statistic for the multivariate survival model (1).

3.2 Proposed Influence Statistic for Multivariate Survival Data

From Sect. 1.1, it is noted that the estimation of β in model (1) is completed by a numerical technique. So, manual approximation of $\Delta\widehat{\beta}_i$ can be done by observing changes in one-step Newton–Raphson approximations to $\widehat{\beta}$, after refitting the model to the data for each removal of a cluster. But, this is again time-consuming as stated before, since a model will have to be refitted upon removal of a cluster. An extension of the score residual (18) that results from fitting model (1) to data once has been proposed in Kaombe and Manda (2022), which estimates influence of a cluster on $\widehat{\beta}$ for model (1). The focus in this chapter is on cluster influence on $\widehat{\beta}$ and not $\widehat{b}_i$, because the estimator $\widehat{\beta}$ depends on all data records regardless of clusters, such that dropping a cluster will have an impact on $\widehat{\beta}$. This is not the case with estimators for random effects, $\widehat{b}_i$. Due to assumed independence of clusters, deleting one cluster does not affect estimate of random effect $\widehat{b}_i$ for another cluster (Xiang et al., 2002).

To study influence for grouped observations, we first define a leverage and a residual for a single observation ij at a given time t_{ij}. The score process (5) derived for model (1) is essentially a row vector of differences between the individual ij covariate value and the average for the covariates of all individuals at risk at time t_{ij}. In essence, this is analogous to leverage in linear models (Sarkar et al., 2011; Zhang, 2016). For individual ij, we let $r_{ij} = exp(X_{ij}^T\widehat{\beta} + \widehat{b}_i)$ be its risk score. Then, at the il-th event-time t_{il}, the Schoenfeld residual (or leverage) (Schoenfeld, 1982), denoted by w_{il}, is given by:

$$\begin{aligned} w_{il} &= X_{il} - \frac{\sum_{s\in R(t_{il})} r_{is}X_{is}}{\sum_{s\in R(t_{il})} r_{is}}, \\ &= X_{il} - \bar{X}(\widehat{\beta}, \widehat{b}_i, t_{il}), \end{aligned} \tag{19}$$

where $r_{is} = exp(X_{is}^T\widehat{\beta} + \widehat{b}_i)$ is the risk score for unit ij in the risk set $R(t_{il})$, and X_{il} is the covariate vector of the individual experiencing the event at time t_{il}. Further, $\widehat{\beta}$ and $\widehat{b}_i$ are, respectively, fixed and random effects terms estimated from the log-likelihood (4). In addition, $\bar{X}(.)$ is a vector whose elements are the conditional weighted means of the covariates values for the individuals at risk of event at time t_{ij}. Hence, the dimension of (20) is $1 \times p$ vector corresponding to each ij-th unit in the risk set (Kaombe & Manda, 2021).

The quantity (20) is also a residual proposed by Schoenfeld (1982) that sums the score processes (5) of units with failure times at each unique event, assuming no ties. Denote $\mathbf{W}_{il}$ as leverages w_{il} for all n_l data points in the risk set and with p covariates, then $\mathbf{W}_{il}$ will be $n_l \times p$ matrix. Furthermore, $w_{il} \in [-\infty, +\infty]$, with mean $E(w_{il}) = E(X_{il}) - E[\bar{X}(\widehat{\beta}, \widehat{b}_i, t_{il})] = E(X_{il}) - E(X_{il}) = 0$. The value 0 of w_{il} corresponds to observations with intermediate covariates values and are thus close to the weighted average for covariate X_{il}, and hence their leverage on the fitted survival curve is negligible. While large negative and large positive values of w_{il} correspond to observations that have unusual covariates values that are far from the weighted average of X_{il}, and hence they have high leverage on the fitted survival curve (Kaombe & Manda, 2021).

A residual, on the other hand, means the difference between the observed and fitted outcomes. The smaller this is, the better the model's fit for the observation of interest (Aguinis et al., 2013; Zhang, 2016). For survival data, one of the residuals is the martingale, defined in Eq. (11), which is an estimate of difference in counts of observed and estimated number of events at each observation time (Therneau et al., 1990). The extension of martingale residual for clustered survival model (1) is defined in Eq. (11) as a $n_l \times 1$ stacked vector of residuals for units in the risk set $R(t_{il})$ given by:

$$m(t_{il}) = N(t_{il}) - \widehat{H}_0(t)exp(X_{il}^T\widehat{\beta} + \widehat{b}_i)$$

$$\Rightarrow \begin{bmatrix} m(t_{11}) \\ \vdots \\ m(t_{1n_1}) \\ m(t_{21}) \\ \vdots \\ m(t_{2n_2}) \\ \vdots \\ m(t_{M1}) \\ \vdots \\ m(t_{Mn_M}) \end{bmatrix} = \begin{bmatrix} N(t_{11}) - \widehat{H}_0(t)exp(X_{11}^T\widehat{\beta} + \widehat{b}_1) \\ \vdots \\ N(t_{1n_1}) - \widehat{H}_0(t)exp(X_{1n_1}^T\widehat{\beta} + \widehat{b}_1) \\ N(t_{21}) - \widehat{H}_0(t)exp(X_{21}^T\widehat{\beta} + \widehat{b}_2) \\ \vdots \\ N(t_{2n_2}) - \widehat{H}_0(t)exp(X_{2n_2}^T\widehat{\beta} + \widehat{b}_2) \\ \vdots \\ N(t_{M1}) - \widehat{H}_0(t)exp(X_{M1}^T\widehat{\beta} + \widehat{b}_M) \\ \vdots \\ N(t_{Mn_M}) - \widehat{H}_0(t)exp(X_{Mn_M}^T\widehat{\beta} + \widehat{b}_M) \end{bmatrix}, \tag{20}$$

where $\widehat{H}_0(t) = \int_{-\infty}^{t} h_0(s)ds$ is the estimated cumulative baseline hazard. The residual (21) has values in the range $(-\infty, 1]$, because $N(t_{il})$ is either 0 or 1 and $\widehat{H}_0(t)exp(X_{il}^T\widehat{\beta} + \widehat{b}_i)$ has values in the interval $[0, \infty)$. In addition, $E(m(t_{il})) = E(N(t_{il})) - E(\widehat{H}_0(t)exp(X_{il}^T\widehat{\beta} + \widehat{b}_i)) = E(N(t_{il})) - E(N(t_{il})) = 0$, since the off-minus quantity in (21) is the average number of events.

Both leverage quantity (20) and residual (21) have correlated values for subjects that are in the same cluster due to shared random effect, but independent values between clusters. Due to this property, we utilise the independence of clusters to derive an influence statistic for detecting impact of dropping a cluster on the estimate of β. Influence of an observation on regression parameter estimates is a product of its outlier and leverage values. Many studies, for example, Cook (1977) for linear models, Zewotir and Galpin (2005) for linear mixed-effects models, Therneau et al. (1990) for univariate survival models, have shown this. Thus, in deriving influence statistics, appropriate case-deletion residual and leverage measures need to be defined first. Using the residual defined in (21) and leverage in (20) for model (1), we propose an analogue of the score residual (18) (Kaombe & Manda, 2021) to measure influence of a cluster on $\widehat{\beta}$ for the model (1) as a vector product of values of vector (21) and those of columns of matrix (20) for subjects under risk set $R(t_{il})$ in the same cluster i, given by:

$$v_i(\widehat{\beta}) = [m(t_{il})]^T \times \mathbf{W}_{il}. \tag{21}$$

The extended score residual (22) is an $((1 \times n_1) \times (n_1 \times p) \ldots (1 \times n_M) \times (n_M \times p)) = M \times p$ matrix, as the value $v_1(\widehat{\beta})$ for first cluster will be a $(1 \times n_1) \times (n_1 \times p) = 1 \times p$ vector reflecting influence of first cluster on each $\widehat{\beta}$ for p covariates, while $v_2(\widehat{\beta})$ for second cluster will be a $(1 \times n_2) \times (n_2 \times p) = 1 \times p$ vector, and so forth. The measure (22) will quantify joint influence of observations in a cluster on the estimate $\widehat{\beta}$, since each of its components is a measure of joint extremity of cluster observations in terms of fitted survival outcomes, as well as in covariates' values

off the fitted survival curve. Since $\mathbf{W}_{il}$ in (22) has elements $w_{il} \in [-\infty, +\infty]$ and $m(t_{il}) \in (-\infty, 1]$, both with mean 0, then the proposed influence statistic (22) is expected to have mean 0. Large positive value of the proposed statistic (22) means a cluster has majority of subjects that have high positive values in w_{il} that coincide with high positive values in $m(t_{il})$, or large negative values in w_{il} coinciding with large negative values in $m(t_{il})$. Technically, this means the cluster has majority of large positive leverage subjects that experienced more events (i.e. failed too early) than predicted by the model or has most subjects with large negative leverage that survived longer than predicted by the model. Hence, such a cluster requires further investigation (Kaombe & Manda, 2021).

On the other hand, large negative value of (22) implies that a cluster has majority of subjects that have large positive leverage w_{il} that coincide with large negative values of the residual $m(t_{il})$ or vice versa. In other words, this implies that the cluster has majority of large positive leverage observations that experienced fewer events (i.e. survived longer) than predicted by the model or has majority of large negative leverage subjects that failed too early than predicted by the model. Again, such a cluster will need further investigation. The values of (22) that are close to zero imply most subjects of the corresponding clusters have either leverage close to zero or residual close to zero, hence such clusters have no issues for follow-up investigation. To decide on influential groups, some studies in linear mixed-effects models have used a cut-off of $\pm 2/\sqrt{M}$ for the values of the influence statistic (Belsley et al., 2005; Nieuwenhuis et al., 2012). However, graphical methods or relative comparisons of influence values for groups are commonly used (Zewotir & Galpin, 2007). We applied graphical techniques in the next two sections to examine influential clusters to the fitted survival mixed models using the proposed influence statistic (22) (Kaombe & Manda, 2021).

3.3 Numerical Example

The data generated from a simulation study described in Sect. 2.2 were utilised to evaluate performance of the proposed influence statistic developed in Sect. 3. The two clusters in which the generated survival data from model (16) involved perturbed β_1 and β_2 were subjected to examination to observe whether they would be identified by the proposed influence statistic (22). The same assessment criterion described in Sect. 2.2 was used, that is, through percentage of simulations for which the proposed influence statistic correctly identified the two target clusters as having influence on β_1 or β_2 using similar cutoffs as described in Sect. 2.2. Upon fitting the model to the simulated data, the proposed influence statistic was computed and its performance evaluated.

3.3.1 Results of Simulations

Like in Sect. 2.2, we inspected performance of the derived influence statistic through graphical techniques prior to its detailed evaluation. The findings in Fig. 3 for two scenarios of simulations showed that the proposed influence statistic had detected influence of clusters 1 and 2 on the estimates of $\widehat{\beta}_2$ and $\widehat{\beta}_1$, respectively. The values of the statistic were outstandingly higher in the first two clusters, where coefficients of the data-generating model were perturbed, than in the other clusters. This study therefore assessed success rates of the proposed influence statistic under each simulation scenario using the cutoffs described before.

Table 5 provides success rates of the proposed influence statistic in detecting impact of cluster 1 or 2 on $\widehat{\beta}_1$ over 100 and 1000 simulations. The results show that the statistic correctly identified the two influential clusters with high percentage, when the perturbations involved β_1 or β_1 and β_2 jointly. The rates for influence of cluster 1 or 2 on $\widehat{\beta}_1$ were relatively low, when it was β_2 that was perturbed. The results also show that the sensitivity of the proposed residual improved with cluster sample size, such that the success rates were as high as 100% where cluster size was 500 and lower with varying degrees when cluster size was 80 subjects. In addition, performance of the statistic improved with size of perturbed parameter value, and this was noticeable where cluster sample sizes were low.

It is also shown that performance of the influence statistic was not different between 100 and 1000 simulation sizes, when cluster sample size was 500 subjects. But the success rates generally slumped in 1000 replications, when cluster size was 80. Finally, the results show that the influence statistic was equally effective across different number of clusters per dataset.

The results in Table 6 are for success rates over 100 and 1000 replications for the proposed influence statistic in identifying cluster 1 or 2 as having influence on

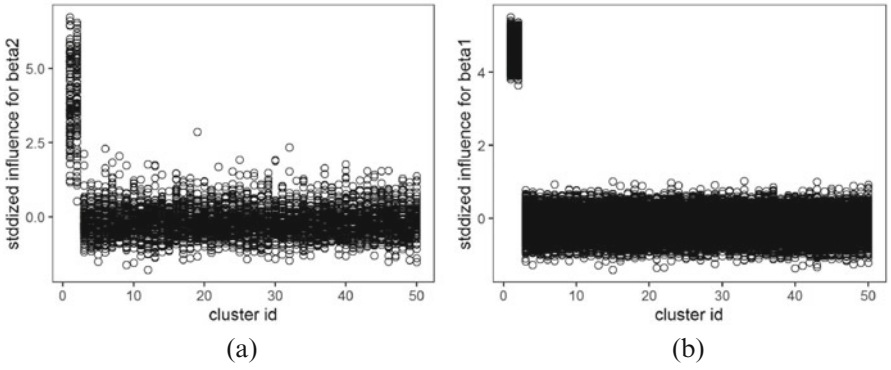


Fig. 3 Plots of cluster influence on $\widehat{\beta}_1$ or $\widehat{\beta}_2$ under two cases of simulations. (**a**) Scatter plots of influence statistic vs cluster id for a case of data with perturbed $\beta_2 = 2.0$ in 2 of 50-clusters sample, each with 80 subjects and with 100 replications. (**b**) Scatter plots of influence statistic vs cluster id for a case of data with perturbed $\beta_1 = 2.7$ in 2 of 50-clusters sample, each with 500 subjects and with 1000 replications. Source: Researcher

Table 5 Percentage of simulations[a] that identified cluster 1 or 2 as influential to $\widehat{\beta}_1$

				100 replicates		1000 replicates	
M	n_i	β_1	β_2	%Cluster1	%Cluster2	%Cluster1	%Cluster2
10	80	1.8	1	84	87	60.9	59.4
	80	2.7	1	100	100	99.3	99.2
10	500	1.8	1	100	100	100	100
	500	2.7	1	100	100	100	100
20	80	1.8	1	74	75	46.4	44.9
	80	2.7	1	99	99	95.3	95.1
20	500	1.8	1	100	100	100	100
	500	2.7	1	100	100	100	100
50	80	1.8	1	34	31	10.7	11.8
	80	2.7	1	75	75	52.7	55.3
50	500	1.8	1	100	100	100	100
	500	2.7	1	100	100	100	100
10	80	0.5	2.0	19	22	42	49
	80	0.5	2.5	36	38	32.3	36.3
10	500	0.5	2.0	27	29	13.1	15.1
	500	0.5	2.5	47	39	41.1	38.5
20	80	0.5	2.0	27	25	10.6	13.1
	80	0.5	2.5	27	31	36.9	40.4
20	500	0.5	2.0	29	30	18.9	20.4
	500	0.5	2.5	60	51	43.9	45
50	80	0.5	2.0	30	29	13.2	12.9
	80	0.5	2.5	60	54	43.7	42.8
50	500	0.5	2.0	30	28	23.5	22.1
	500	0.5	2.5	63	62	47.6	48.6
10	80	1.8	2.0	69	77	57.5	59
	80	2.7	2.5	99	96	84.6	83
10	500	1.8	2.0	100	100	100	100
	500	2.7	2.5	100	100	100	100
20	80	1.8	2.0	70	69	43.8	46.2
	80	2.7	2.5	92	92	76.6	74.7
20	500	1.8	2.0	100	100	100	100
	500	2.7	2.5	100	100	100	100
50	80	1.8	2.0	67	51	45.8	44.9
	80	2.7	2.5	86	87	71.9	70.6
50	500	1.8	2.0	100	100	100	100
	500	2.7	2.5	100	100	100	100

[a] No perturbations were done to data in other clusters than 1 and 2; in those other clusters, model (16) had $\beta_1 = 0.5$ and $\beta_2 = 1$

$\widehat{\beta}_2$. The findings show that the proposed influence statistic highly detected impact of first two clusters on $\widehat{\beta}_2$, when it was β_2 or jointly β_2 and β_1 that was perturbed during data generation. The success rates of the statistic in detecting influence of cluster 1 or 2 on $\widehat{\beta}_2$ were low when it was β_1 that was perturbed.

As was the case with $\widehat{\beta}_1$, the success rates of the statistic in detecting impact of cluster 1 or 2 on $\widehat{\beta}_2$ improved with cluster size, as the rates were consistently high for cluster sizes of 500 and low with cluster sizes of 80 subjects. Again, the performance of the statistic improved with size of perturbed parameter value, a situation that was also noticeable in low cluster sizes like before. Likewise, there was no difference in performance of the statistic between 100 and 1000 simulation sizes, this was much apparent in large cluster sample sizes. Lastly, it is also shown that the influence statistic performed equally across different number of clusters per sample.

3.4 Application of the Influence Statistic on Malawi Child Survival Data

Using the child survival data and the model described in Sect. 2.3, we present application of the proposed influence statistic (22) as demonstrated in Kaombe and Manda (2021). The application tracks how the proposed influence statistic would estimate influence of dropping each of the 56 subdistricts on effect of being female on under-five mortality. This is for demonstration's sake, otherwise one could also assess influence of the subdistricts on the other covariates in the model. Once again, national under-five mortality rate of 63 deaths per 1000 live births (Malawi National Statistical Office (NSO) and ICF, 2017) was used as baseline hazard during computation of influence values.

3.4.1 Results for Influential Subdistricts on Effect of Being Female on Under-Five Mortality

The results in Fig. 4, showed that the proposed influence statistic detected *Kasungu* rural as having positive influence on effect of female gender on child mortality. This means that *Kasungu* rural subdistrict had majority of children with high leverage on estimated mortality that had also died too early than predicted by the model, such that dropping this cluster from the model would cause a significant change on estimated effect of female gender on child mortality. The statistic also detected *Karonga* rural, *Mzimba* rural, *Dowa* urban, *Thyolo* urban, and *Phalombe* urban as having negative influence on effect of being female on child mortality. This implies that these subdistricts had majority of children with high leverage on estimated mortality who had also survived longer than predicted by the model, such that removing each of these clusters from analysis would impact on estimated effect of female gender on child mortality.

Table 6 Percentage of simulations[a] that identified cluster 1 or 2 as influential to $\widehat{\beta}_2$

				100 replicates		1000 replicates	
M	n_i	β_1	β_2	%Cluster1	%Cluster2	%Cluster1	%Cluster2
10	80	1.8	1	2	2	0.9	0.7
	80	2.7	1	4	4	1.2	1.3
10	500	1.8	1	0	0	0	0
	500	2.7	1	0	0	2.6	2.3
20	80	1.8	1	14	12	4.8	5.5
	80	2.7	1	0.8	0.6	4.6	4.6
20	500	1.8	1	0.9	1.2	1.3	1.4
	500	2.7	1	1	0.8	2	1.2
50	80	1.8	1	34	40	13.7	14.8
	80	2.7	1	34	33	19.6	20.0
50	500	1.8	1	26	18	13.4	11
	500	2.7	1	18	14	8.5	7.4
10	80	0.5	2.0	94	97	93.8	93.5
	80	0.5	2.5	98	100	98.5	97.9
10	500	0.5	2.0	100	100	100	100
	500	0.5	2.5	100	100	100	100
20	80	0.5	2.0	98	98	93.4	92.7
	80	0.5	2.5	100	100	97.7	97.4
20	500	0.5	2.0	100	100	100	100
	500	0.5	2.5	100	100	100	100
50	80	0.5	2.0	99	97	94.4	94.6
	80	0.5	2.5	100	100	97.3	97.6
50	500	0.5	2.0	100	100	100	100
	500	0.5	2.5	100	100	100	100
10	80	1.8	2.0	72	77	39.1	42.5
	80	2.7	2.5	99	92	81.6	81.3
10	500	1.8	2.0	100	100	100	100
	500	2.7	2.5	100	100	100	100
20	80	1.8	2.0	64	73	46.7	45
	80	2.7	2.5	88	81	65.2	63.7
20	500	1.8	2.0	100	100	100	100
	500	2.7	2.5	100	100	100	100
50	80	1.8	2.0	59	53	43.4	43.4
	80	2.7	2.5	78	74	63.6	61.5
50	500	1.8	2.0	100	100	99.9	99.8
	500	2.7	2.5	100	100	100	100

[a] No perturbations were done to data in other clusters than 1 and 2; in those other clusters, model (16) had $\beta_1 = 0.5$ and $\beta_2 = 1$

3.4.2 Impact of the Identified Influential Subdistricts on Estimate of Effect of Being Female on Under-Five Mortality

Upon identifying the influential subdistricts on estimate of effect of female gender in the model, we used the manual approach to analyse their impact on the estimates, that is, we observed the changes in estimates upon refitting the model to data without the detected subdistricts. This was demonstrated with two extreme influential subdistricts: one with positive influence and the second with negative influence. The results are in Table 7. It was shown that removal of *Kasungu* rural cluster from analysis resulted in further reduction in logarithm of hazard of death for female children by 0.0187. Thus, the survival model was better off without data from *Kasungu* rural cluster. This was also noticed with the reduction in p-value by 0.001. While, dropping *Mzimba* rural cluster increased the logarithm of hazard of death in female children by 0.0053. Thus, the data from *Mzimba* rural cluster were required in the model. This is also reflected in the p-value that got higher upon removing this cluster. Removing both clusters from analysis resulted in reduction of logarithm of hazard of death, but not as much as when *Kasungu* rural cluster was dropped alone.

Thus, the effect of dropping the two clusters at the same time did not add value to the estimation compared to dropping each one of them separately. This was the case since *Kasungu* rural cluster had positive influence while *Mzimba* rural negative influence on estimate of effect of female gender on child mortality, so that their combined influence could not be in either direction but could just mask the influence of either of the two. The standard errors of the parameter estimates slightly increased in each case, implying that the original estimates from full data were biased. The variance of random effects also got lower in both cases. Further, the results vindicate the magnitude of influence of each of the two clusters as reported by our proposed statistic in the previous paragraph. It is shown in Table 7 that impact of either cluster on the estimate of effect of female gender on mortality was about the same magnitude either side of the estimate obtained using full dataset, which agrees with findings from the proposed statistic in Fig. 4.

4 Conclusion

In this chapter, we have developed outlier and influence statistics for analysing multivariate survival data model, by extending methods developed for linear mixed-effects and univariate survival models. Simulation studies have shown that the proposed statistics correctly identified the outlying and influential clusters in the analysis of multivariate survival data, over 99% and 98% of the time, respectively. The performance of the statistics improved with cluster sample size. The proposed influence statistic detects both direction and magnitude of influence of a cluster on regression parameter estimates. The application of the proposed statistics should be done through relative comparisons of values of the statistics across clusters, with large positive values of the outlier statistic signalling outlying clusters; and large

Table 7 Estimates for effect of being female on child mortality with and without Kasungu rural or Mzimba rural clusters or both in the Malawi child survival dataset, for a frailty Cox model

Parameter	Full dataset	Without Kasungu rural (diff[a])	Without Mzimba rural (diff[a])	Without both (diff[a])
$\widehat{\beta}$	−0.2343	−0.253 (0.0187)	−0.229 (−0.0053)	−0.2475 (0.0132)
$se(\widehat{\beta})$	0.0869	0.088 (−0.0011)	0.088 (−0.0011)	0.0892 (−0.0023)
$p\text{-}value$	0.0070	0.0041 (0.0029)	0.0094 (−0.0024)	0.0055 (0.0015)
$var(re)$	0.0257	0.0236 (0.0021)	0.026 (−0.0003)	0.0237 (0.0020)

[a] diff = parameter estimate under full dataset−parameter estimate under reduced data, $se(\widehat{\beta})$ is standard error of $\widehat{\beta}$, $var(re)$ is variance of random effects

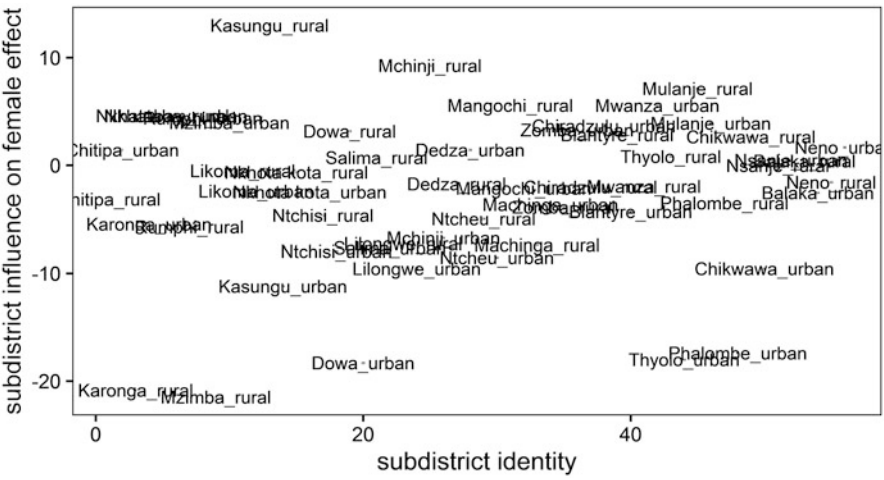


Fig. 4 Subdistrict level estimates of the proposed influence statistic for effect of female gender upon fitting a frailty Cox hazard regression model to Malawi child survival data, 2015–16 MDHS. Source: Researcher

positive and large negative values of the influence statistic, respectively, indicating positively and negatively influential clusters on particular fixed regression effect (Zewotir & Galpin, 2006).

With a frailty survival model, where observations in a cluster share random effect, it was observed that outlying tendency of a cluster, based on our method, is largely influenced by fixed covariates rather than random covariates. Large sample size for a statistic leads to lowered uncertainty in repeated sampling (Hemez et al., 2010), which is why performance of our proposed statistics improved with sample size. The application showed that outlier detection methods based on inspection of single observations' residuals as done in linear mixed models (Langford & Lewis, 1998) proved difficult to detect group outliers for multivariate survival model. The dataset had high censoring rate of 95%, which usually cause deviance residuals to lose symmetry, especially from 40% and above (Therneau et al., 1990). This might have affected relevance of outlier methods based on residual inspection of the data, as the methods are applied with an assumption that residuals are symmetric about mean zero. The original data revealed that the identified outlying subdistricts from urban locations had lower under-five mortality rates, while those that were rural-based had higher rates. Studies have attributed low child mortality in urban settings due to high access to healthcare services by children, resulting from short distances travelled by their mothers to clinics, compared to the children from rural locations (Ustrup et al., 2014).

Finally, we recommend that an analysis of multivariate survival data should be accompanied by an assessment of outlying and influential clusters to avoid having biased estimates and inaccurate conclusions. For the identified influential clusters, one could investigate contribution of individual units in making the

clusters as such (Xiang et al., 2002; Zewotir & Galpin, 2006). We recommend such analyses for future research. The outlier statistics hinge on defining relevant structure of a residual for the model at hand. Future studies could develop outlier statistics for various formulations of the multivariate survival model, other than time-constant covariates model, like stratified and time-dependent survival model. We also recommend an in-depth study to investigate child healthcare practices in the identified outlying or influential subdistricts regarding child survival in Malawi.

Acknowledgments This work was supported through the DELTAS Africa Initiative SSACAB (Grant No.107754/Z/15/Z). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences' (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the authors and not necessarily those of the AAS, NEPAD Agency, Wellcome Trust or the UK government. We are also thankful to South African Medical Research Council, in Pretoria, South Africa, for hosting the main author in this research during his research visit to the co-author at MRC, who was also his PhD mentor, in the year 2019.

Appendices

R Code for Applying Derived Outlier Statistic on Child Survival Data

```
rm(list=ls())
library(foreign)
library(survival)
args(coxph)
library(ggplot2)
library(ggrepel)
library(data.table)
library(dplyr)
library(readstata13)


mydata = read.dta("c:/Users/User/Desktop/applicationI/DHSrdcd.
dta",convert.factors=T)
tmo<- Sys.time()
ntimes <-data.frame(mydata %>%count(subdistrict))$n

ntimes

for(k in 1:1)
{
        model <- coxph(Surv(as.numeric(agedth),as.numeric(died))
~Sex+bordcat+bwgtcat+pbirth+frailty(subdistrict, distribution
="gaussian",sparse=F,method="reml"),data=mydata)
```

```
        for (j in 1:56)
        {
                dt <- data.frame(cbind(newDist=1:56,coefSex=rep
(coef(model)[1],56),coefbordcat=rep(coef(model)[2],56),
coefbweight=rep(coef(model)[3],56),coefpbirth=rep(coef
(model)[4],56),randeffect=coef(model)[-4:-1]))
                dt2 <- as.data.frame(dt[rep(1:nrow(dt),ntimes),])
        }

        dt2$martingale <- mydata$died - (0.063*mydata$agedth*exp(
         mydata$Sex*dt2$coefSex+mydata$bordcat*dt2$coefbordcat
     +mydata$bwgtcat*dt2$coefbweight+mydata$pbirth*dt2$
     coefpbirth+dt2$randeffect))
        dt2$sign=ifelse(dt2$martingale>0,1,-1)
        dt2$deviance <- dt2$sign*dt2$martingale*sqrt(-2*(dt2
    $martingale+mydata$died*log(mydata$died-dt2$martingale)))
        dt2$grandmean <- setDT(dt2)[,lapply(.SD,mean,na.rm=TRUE),
    .SDcols="deviance"]

        mydata4 <- data.frame(cbind(mydata,dt2))

        write.dta(mydata4, paste0("c:/Users/User/Desktop/
␣␣␣␣applicationIIII/mydata4.dta"))
}
tm1<- Sys.time()
tm1 - tmo

outliermat <- matrix(NA,nrow =56,ncol = 6)
outliermat8 <- data.frame(outliermat)
colnames(outliermat8) <- c("ID","meanclusdev","wtngrpVar",
"grandavg","btwngrpVar","rationvar")

tmo<- Sys.time()

for(k in 1:1)
{

        outliermat8[,1]<- 1:56
        outliermat8[,2] <- setDT(mydata4)[,lapply(.SD,mean,na.
                        rm=TRUE),by=subdistrict,.SDcols="
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣deviance"][,2]
        outliermat8[,3]<- setDT(mydata4)[,lapply(.SD,var,na.
                        rm=TRUE),by=subdistrict,.SDcols=
                        "deviance"][,2]
        outliermat8[,4]<- setDT(mydata4)[,lapply(.SD,mean,na.
                        rm=TRUE),by=subdistrict,.SDcols=
                        "grandmean"][,2]
        outliermat8[,5] <- sum(ntimes*(outliermat8$meanclusdev
                         - outliermat8$grandavg))^{2}/(56-1)
        outliermat8[,6]<- outliermat8$wtngrpVar/outliermat8$
                        btwngrpVar

        outliermat8 = cbind.data.frame(outliermat8)
```

```
}

write.dta(outliermat8, paste0("c:/Users/User/Desktop/applicationI
/outlierd8.dta"))

tm1<- Sys.time()
tm1 - tmo
```

R Code for Applying Derived Influence Statistic on Child Survival Data

```
rm(list=ls())
library(survival)
args(coxph)
library(foreign)
library(ggplot2)
library(ggrepel)
library(dplyr)
library(data.table)
library(readstata13)

ourdata = read.dta("C:/Users/User/Desktop/izi/influence2.dta",
convert.factors=F)
tmo<- Sys.time()
ntimes <-data.frame(ourdata %>%count(v023))$n

for(k in 1:n)
{
        model <- coxph(Surv(as.numeric(agedth),as.numeric(died))
~Sex+bordcat+bwgtcat+pbirth+frailty(v023, distribution=
"gaussian",sparse=F,method="reml"),data=ourdata)

        for (j in 1:56)
        {
                dt <- data.frame(cbind(newDist=1:56,coefSex=rep
        (coef(model)[1],56),coefbordcat=rep(coef(model)
        [2],56),coefbweight=rep(coef(model)[3],56),
        coefpbirth=rep(coef(model)[4],56),randeffect=coef
        (model)[-4:-1]))
                dt2 <- as.data.frame(dt[rep(1:nrow(dt),ntimes),])
        }

        dt2$martingale <- ourdata$died-(0.063*ourdata$agedth*exp(
         ourdata$Sex*dt2$coefSex+ourdata$bordcat*dt2$coefbordcat
    +ourdata$bwgtcat*dt2$coefbweight+ourdata$pbirth*dt2$
    coefpbirth+dt2$randeffect))

        ourdata <- data.frame(cbind(ourdata,dt2))

        ourdata$numerator_Sex <- ourdata$Sex* exp(ourdata$Sex*
ourdata$coefSex+ourdata$bordcat*ourdata$coefbordcat+ourdata$
bwgtcat*ourdata$coefbweight+ourdata$pbirth*ourdata$coefpbirth
```

```
+ourdata$randeffect)

        ourdata$numerator_bord <- ourdata$bordcat * exp(ourdata$
Sex*ourdata$coefSex+ourdata$bordcat*ourdata$coefbordcat+
ourdata$bwgtcat*ourdata$coefbweight+ourdata$pbirth*ourdata$
coefpbirth+ourdata$randeffect)

        ourdata$numerator_bweight <- ourdata$bwgtcat * exp(ourdata
$Sex*ourdata$coefSex+ourdata$bordcat*ourdata$coefbordcat+
ourdata$bwgtcat*ourdata$coefbweight+ourdata$pbirth*ourdata$
coefpbirth+ourdata$randeffect)

        ourdata$numerator_pbirth <- ourdata$pbirth * exp(ourdata$
Sex*ourdata$coefSex+ourdata$bordcat*ourdata$coefbordcat+
ourdata$bwgtcat*ourdata$coefbweight+ourdata$pbirth*ourdata$
coefpbirth+ourdata$randeffect)

        ourdata$denominator_Sex <- exp(ourdata$Sex*ourdata$coefSex
+ourdata$bordcat*ourdata$coefbordcat+ourdata$bwgtcat*ourdata$
coefbweight+ourdata$pbirth*ourdata$coefpbirth+ourdata$randeffect)

        ourdata$sum_numSex <- setDT(ourdata)[,lapply(.SD,sum,na.
rm=TRUE),by=v023,.SDcols="numerator_Sex"][,2][rep(1:nrow(dt),
ntimes),]
        ourdata$sum_numbord <- setDT(ourdata)[,lapply(.SD,sum,na.
rm=TRUE),by=v023,.SDcols="numerator_bord"][,2][rep(1:nrow(dt),
ntimes),]
        ourdata$sum_numbweight <- setDT(ourdata)[,lapply(.SD,sum,
na.rm=TRUE),by=v023,.SDcols="numerator_bweight"][,2][rep(1:nrow
(dt),ntimes),]
        ourdata$sum_numpbirth <- setDT(ourdata)[,lapply(.SD,sum,na.
rm=TRUE),by=v023,.SDcols="numerator_pbirth"][,2][rep(1:nrow(dt),
ntimes),]
        ourdata$sum_denSex <- setDT(ourdata)[,lapply(.SD,sum,na.
rm=TRUE),by=v023,.SDcols="denominator_Sex"][,2][rep(1:nrow(dt),
ntimes),]

        ourdata$leverage_Sex <- ourdata$Sex - (ourdata$sum_numSex
/ourdata$sum_denSex)
        ourdata$leverage_bord <- ourdata$bord - (ourdata$sum_
numbord/ourdata$sum_denSex)
        ourdata$leverage_bweight <- ourdata$bwgtcat - (ourdata$
sum_numbweight/ourdata$sum_denSex)
        ourdata$leverage_pbirth <- ourdata$pbirth - (ourdata$
sum_numpbirth/ourdata$sum_denSex)

        ourdata$scoresdSex <- ourdata$martingale * ourdata$
leverage_Sex
        ourdata$scoresdbord <- ourdata$martingale * ourdata
$leverage_bord
        ourdata$scoresdbweight <- ourdata$martingale *
ourdata$leverage_bweight
        ourdata$scoresdpbirth <- ourdata$martingale *
ourdata$leverage_pbirth
```

```
        write.dta(ourdata,file = "data4.dta")

}
tm1<- Sys.time()
tm1 - tmo

# B. Computing group score residual

influmat <- matrix(NA,nrow =56,ncol =5)
influmat <- data.frame(influmat)
colnames(influmat) <- c("ID","Influ_Sex","Influ_bord","Influ_
bweight","Influ_pbirth")

tmo<- Sys.time()
influmat_all = matrix(NA,nrow =56,ncol =5)
colnames(influmat_all) <- c("ID","Influ_Sex","Influ_bord","Influ_
bweight","Influ_pbirth")
for(k in 1:n)
{
        influmat[,1]<- 1:56
        influmat[,2] <-setDT(ourdata)[,lapply(.SD,sum,na.rm=TRUE),
by=v023,.SDcols="scoresdSex"][,2]
        influmat[,3] <-setDT(ourdata)[,lapply(.SD,sum,na.rm=TRUE),
by=v023,.SDcols="scoresdbord"][,2]
        influmat[,4] <-setDT(ourdata)[,lapply(.SD,sum,na.rm=TRUE),
by=v023,.SDcols="scoresdbweight"][,2]
        influmat[,5] <-setDT(ourdata)[,lapply(.SD,sum,na.rm=TRUE),
by=v023,.SDcols="scoresdpbirth"][,2]
}
write.dta(influmat,file = "influence2.dta")

tm1<- Sys.time()
tm1 - tmo
```

References

Abrahantes, J., & Burzykowski, T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal, 47*(6), 847–862.

Aguinis, H., Gottfredson, R., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270–301.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, vol. 571. New York: John Wiley & Sons.

Brilleman, S., Rory, W., Moreno-Betancur, M., & Crowther, M. (2018). simsurv: A package for simulating simple or complex survival data. In *UseR! Conference 2018, Brisbane, Australia*. Monash University.

Cain, K., & Lange, N. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics, 40*(2), 493–499.

Cook, D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*(1), 15–18.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–202.

Das, M., & Gogoi, B. (2015). Influential observations and cutoffs of different influence measures in multiple linear regression. *International Journal of Computational and Theoretical Statistics, 2*(2), 79–85.

Ha, I., Sylvester, R., Legrand, C., & MacKenzie, G. (2011). Frailty modelling for survival data from multi-centre clinical trials. *Statistics in Medicine, 30*(17), 2144–2159.

Hemez, F., Atamturktur, H. S., & Unal, C. (2010). Defining predictive maturity for validated numerical simulations. *Computers & Structures, 88*(7–8), 497–505.

Kaombe, T. M., & Manda, S. O. M. (2021). Detecting influential data in multivariate survival models. *Communications in Statistics – Theory and Methods*. https://doi.org/10.1080/03610926.2021.1982983.

Kaombe, T. M., & Manda, S. O. M. (2022). A novel outlier statistic in multivariate survival models and its application to identify unusual under-five mortality sub-districts in Malawi. *Journal of Applied Statistics*. https://doi.org/10.1080/02664763.2022.2043255.

Langford, I., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 161*(2), 121–160.

Liang, K., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual Review of Public Health, 14*(1), 43–68.

Maia, R., Madsen, P., & Labouriau, R. (2014). Multivariate survival mixed models for genetic analysis of longevity traits. *Journal of Applied Statistics, 41*(6), 1286–1306.

Malawi National Statistical Office (NSO). (2019). *2018 Malawi population and housing census: main report*. Zomba: Author.

Malawi National Statistical Office (NSO) and ICF. (2017). *2015–16 Malawi demographic and health survey: Key findings*. Zomba, Malawi, and Rockville, Maryland, USA: Author.

Manda, S. (2001). A comparison of methods for analysing a nested frailty model to child survival in Malawi. *Australian & New Zealand Journal of Statistics, 43*(1), 7–16.

Manda, S. (2011). A nonparametric frailty model for clustered survival data. *Communications in Statistics-Theory and Methods, 40*(5), 863–875.

Moriña, D., & Navarro, A. (2014). The R package survsim for the simulation of simple and complex survival data. *Journal of Statistical Software, 59*(2), 1–20.

Nieuwenhuis, R., te Grotenhuis, H., & Pelzer, B. (2012). Influence. me: tools for detecting influential data in mixed effects models: Retrieved from https://repository.ubn.ru.nl/handle/2066/103101.

Nobre, J., & Singer, J. (2011). Leverage analysis for linear mixed models. *Journal of Applied Statistics, 38*(5), 1063–1072.

Ripatti, S., & Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics, 56*(4), 1016–1022.

Sarkar, S., Midi, H., & Rana, S. (2011). Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences, 11*(1), 26–35.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika, 69*(1), 239–241.

Storer, B. E., & Crowley, J. (1985). A diagnostic for cox regression and general conditional likelihoods. *Journal of the American Statistical Association, 80*(389), 139–147.

Therneau, T., Grambsch, P., & Fleming, T. (1990). Martingale-based residuals for survival models. *Biometrika, 77*(1), 147–160.

Ustrup, M., Ngwira, B., Stockman, L., Deming, M., Nyasulu, P., Bowie, C., Msyamboza, K., Meyrowitsch, D., Cunliffe, N., & Bresee, J. (2014). Potential barriers to healthcare in Malawi for under-five children with cough and fever: a national household survey. *Journal of Health, Population, and Nutrition, 32*(1), 68.

Vaida, F., & Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine, 19*(24), 3309–3324.

Van Kempen, G., & Van Vliet, L. (2000). Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry: The Journal of the International Society for Analytical Cytology, 39*(4), 300–305.

Xiang, L., Tse, S.-K., & Lee, A. H. (2002). Influence diagnostics for generalized linear mixed models: applications to clustered data. *Computational Statistics & Data Analysis, 40*(4), 759–774.

Xu, R., Vaida, F., & Harrington, D. (2009). Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. *Statistica Sinica, 19*, 819–842.

Yang, H. (2012). Visual assessment of residual plots in multiple linear regression: A model-based simulation perspective. *Multiple Linear Regression Viewpoints, 38*(2), 24–37.

Zewotir, T. (2008). Multiple cases deletion diagnostics for linear mixed models. *Communications in Statistics-Theory and Methods, 37*(7), 1071–1084.

Zewotir, T., & Galpin, J. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science, 3*(2), 153–177.

Zewotir, T., & Galpin, J. S. (2006). Evaluation of linear mixed model case deletion diagnostic tools by Monte Carlo simulation. *Communications in Statistics-Simulation and Computation, 35*(3), 645–682.

Zewotir, T., & Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed model. *Test, 16*(1), 58–75.

Zhang, Z. (2016). Residuals and regression diagnostics: focusing on logistic regression. *Annals of Translational Medicine, 4*(10), 1–8.

Zhang, D., Sun, J. L., & Pieper, K. (2016). Bivariate mixed effects analysis of clustered data with large cluster sizes. *Statistics in Biosciences, 8*(2), 220–233.

Joint Modelling of Longitudinal and Competing Risks Survival Data

Didjier D. Masangwi, Adamson S. Muula, and Mavuto F. Mukaka

Abstract Biomedical studies may collect longitudinal and survival data in follow-up studies. In randomised controlled trials for malaria treatment, longitudinal parasite count and hemoglobin level and survival outcomes, time to fever resolution or time to parasite clearance, are recorded. The longitudinal and survival data are analysed separately, yet longitudinal outcomes may be important predictors in the survival process. Standard survival analysis methods cannot handle such longitudinal outcomes. In such studies, survival competing risks are possible; thus analysis should consider survival, longitudinal and competing risks. In joint modelling, options for modelling dependence are a key issue as well as choice of random effects distribution. The example used in this work was from sub-Saharan Africa.

Joint modelling framework, mixed-effects models and Cox-specific models for analysis of longitudinal and survival data were applied to malaria dataset from Malawi Liverpool Wellcome Trust. Longitudinal outcomes considered were hemoglobin level and parasite count, while survival outcomes were time to treatment failure due to severe malaria and time to withdrawal (due to adverse effects and protocol violation).

Different survival outcomes observed were severe malaria (4.95%) and withdrawal (10.89%). The longitudinal outcomes were not associated with the risks of

D. D. Masangwi (✉)
Malawi University of Science and Technology, University of Malawi, Chancellor College, Zomba, Malawi

A. S. Muula
University of Malawi, College of Medicine, Africa Center of Excellence in Public Health and Herbal Medicine, Zomba, Malawi
e-mail: amuula@medcol.mw

M. F. Mukaka
Nuffield Department of Medicine, University of Oxford, Oxford, UK

Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand
e-mail: mavuto.mukaka@ndm.ox.ac.uk

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_16

severe malaria and withdrawal in the Cox model. The true hemoglobin level and age were associated with the risk of withdrawal ($p = 0.0111$) and ($p = 0.0305$), respectively, in the joint model, and the separate models were opted to fit the data.

When an association between longitudinal and survival outcomes is of interest, joint models can be considered over separate methods. However, where there is no association, separate models for survival and longitudinal data analysis can be used.

Keywords Survival models · Longitudinal data · Competing risks · Joint models · Severe malaria · Randomised controlled trials · Efficacy · Cox-specific model · Hemoglobin level · Parasite count · Withdrawal · Mixed-effects · Time · Biomedical study · Event · Censoring · Parameter · Random effects · Estimation · BIC · Profile · Covariates · Follow-up · Relative risk · Treatment · Package · Error · Framework · Data · Research · Predicting · Association · Patients

1 Introduction

Biomedical studies may collect repeated measurements of longitudinal data and time to event data during follow-up. A typical example is in AIDS study where CD4 count and viral load are collected longitudinally and time to AIDS or death is also monitored (Elashoff et al., 2008). In cancer studies, longitudinal data such as circulating tumour cells, immune response to a vaccine, a genetic biomarker or a health outcome are recorded (Ibrahim et al., 2010), and during follow-up death or metastasis can occur. In malaria studies, randomised blinded controlled trials are carried out to compare efficacy and safety of drugs and resistance of parasites. During follow-up, one of the measures of interest may be time to fever resolution and time to parasite clearance, with possible longitudinal covariates such as white blood cell count or red blood cell counts and hemoglobin levels.

In order to analyse such data, researchers usually use methods for separate analysis of longitudinal and survival data. In survival analysis, non-parametric and parametric methods such as Weibull, exponential, log-normal and log-logistics are widely applied (Collet, 2003). The effects of covariates on the survival process are modelled using Cox hazard model and extended Cox model for time-dependent covariates, which is theoretically valid for exogenous time-varying covariates but not valid when studying biomarkers (endogenous) and other patient parameters (Andrinopoulou, 2014). The inadequacy comes in as Cox model assumes constancy in the marker's level between time visits. Thus, neglecting these features can lead to underuse of potential variable information which in turn may lead to biased results and conclusion on the effect of the marker.

In cases where there are more than one event failures (competing risks), interpretation of survival probabilities from standard survival models has always been questionable (Kleinbaum & Klein, 2005). However, methods based on cause-

specific Cox model and cumulative incidence curves are used. In survival analysis, censoring is assumed independent of survival time (non-informative).

In longitudinal data, generalised mixed-effects regression models, covariance pattern models, analysis of variance (ANOVA) and generalised estimating equations (GEE) are used to model the repeated measurements (Hedeker & Gibbons, 2006). The impressive aspect of these methods is that of an account for the correlation within the measurements obtained from the same patients/participants and can handle unequally spaced time visits (Andrinopoulou, 2014).

Furthermore, in longitudinal data, missing/incomplete measurements are inevitable. The missing data mechanisms, missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), are commonly encountered in longitudinal studies (Andrinopoulou, 2014). However, selection models, pattern mixture models and shared parameter models for discrete times handle such cases (Molenberghs & Kenward, 2007). MNAR is the most difficult in practice.

Separate analyses of these data do not account for other parameters and the association between the two types of the data. This is the case, because longitudinal covariates can be important predictors of survival process or some other time-to-event. Censoring is assumed noninformative in survival data, yet longitudinal data is affected by informative dropout, especially in cases with competing risks and also the inclusion of time-varying covariates in survival analysis. Ignoring these special features of longitudinal and survival data may lead to underuse of potential variable information and lead to biased results and conclusion (Sudell et al., 2016). There is more work on joint modelling of longitudinal and survival data in literature with single failure event and less work on competing risks (multiple failure events). In this work, competing risks joint models are considered.

We considered the survival methods such as cause-specific hazard model, longitudinal methods such as mixed-effects model and the competing risks joint models that combine the survival and longitudinal processes.

1.1 Competing Risks Joint Models

The joint model is a combination of longitudinal and survival submodels that are linked using an association structure that quantifies the relationship between the outcomes of interest. There is work on joint modelling framework by Rizopoulos (2012), Yu et al. (2004), Tsiatis et al. (1995) and Faucett and Thomas (1996) in recent literature (Andrinopoulou, 2014). In literature, there is considerable work by Williamson et al. (2008) on competing risks joint modelling of longitudinal and survival data, an alternative approach with a different parameterisation. However, in this work, modelling approach by Rizopoulos (2012) was applied to malaria data. We emphasise that in joint modelling, options for modelling dependence is a key issue as well as choice of random effects distribution.

1.1.1 Survival Submodel

As proposed by Rizopoulos (2010), let T_i^* be the true event time for the *i-th* subject, $i = 1,2,..\ n_i$, and let T_i be the observed event time, where $T_i = \min\left(T_i^*, C_i\right)$, where C_i is the censoring time. Let $\delta_i = I\left(T_i^* \leq C_i\right)$, i.e. δ_i is unity for the true event. Let us also assume that $m_i(t)$ is the true or unobserved value of the longitudinal biomarker, a time-dependent covariate measured at different time points t be known. Then $y_i(t)$ is the observed value of the time-dependent covariate at time t, and $y_i(t) = \{y_i(t_{ij}), j = 1, 2, \ldots, n_i\}$. The aim is to associate the true unobserved longitudinal outcome $m_i(t)$ with the hazard of an event.

Assuming we have K different event types; let T_{i1}^*, T_{i2}^*, . . ., T_{iK}^* be the true failure times for K event types. Let T_i be the observed failure time such that $T_i = \min\left(T_{i1}^*, \ldots, T_{iK}^*, C_i\right)$, where C_i is the censoring time. Let δ_i take the values $\{0, 1, 2, \ldots, g\}$, with $\delta_i = 0$ indicating a censored event for subject i and $\delta_i = k$ showing that subject i fails from the *k-th* type of failure, where $k = 1, \ldots, g$. The relative risk model for competing risks is the cause-specific hazard model given as:

$$h_{ik}\left(t|\mathcal{M}_i(t)\right) = h_{0k}(t)\exp\left({\gamma_k}^T w_i + \boldsymbol{\alpha_k}\left(m_i(t)\right)\right) \tag{1}$$

where w_i is vector of baseline covariates, $m_i(t)$ is true value of the longitudinal marker with $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$, $h_{0k}(s)$ is the baseline hazard and $\boldsymbol{\alpha_k}$ quantifies the strength of the association between the marker and the risk of an event/the effect of underlying longitudinal outcome to the risk for an event type k and ${\gamma_k}^T$ is a vector of regression coefficients for baseline covariates. The survival function depends on full history of the marker and is given as:

$$S_{ik}\left(t|\mathcal{M}_i(t)\right) = \exp\left\{-\int_0^t h_{0k}(s)\exp\left({\gamma_k}^T w_i + \boldsymbol{\alpha_k}\left(m_i(s)\right)\right) ds\right\} \tag{2}$$

To avoid misspecification of the underlying parametric distribution of the survival times which in turn leads to under-estimation of standard errors of parameter estimates in the joint model settings, the baseline hazard function is specified (Hsieh et al., 2006). In literature Rizopoulos (2012) proposes a more flexible regression spline model for baseline hazard function in competing events model, given by:

$$\log h_0(t) = k_0 + \sum_d^m k_d B_d\left(t, q\right) \tag{3}$$

where $k^T = (k_0, k_1, \ldots, k_m)$ are the spline coefficients, q denotes the degree of the B-splines basis function B(.) proposed by de Boor (Boor, 1978) and $m = \ddot{m} + q - 1$ where $\ddot{m}$ is the number of interior knots.

1.1.2 Longitudinal Submodel

The relative risk submodel above uses the unobserved longitudinal value $m_i(t)$. In order to determine the effect of the longitudinal outcome on the risk of an event, $m_i(t)$ is estimated, and the complete true longitudinal history $\mathcal{M}_i(t)$ is reconstructed. With continuous longitudinal biomarker, the mixed-effect regression model is given as:

$$y_i(t) = m_i(t) + \varepsilon_i(t) \tag{4}$$

$$= \beta X_i^T(t) + Z_i^T(t)b_i + \varepsilon_i(t), \quad m_i(t) = \beta X_i^T(t) + Z_i^T(t)b_i.$$

where $X_i^T(t)$ is the design vector of fixed coefficients β and $Z_i^T(t)$ is a design vector for random effects b_i. This vector $b_i \sim N(0, D)$ is a latent random variable that can be interpreted as subject-specific effects of $Z_i^T(t)$. D is variance-covariance matrix, and $\varepsilon_i(t)$ is a random error that is assumed to be independent and normally distributed with mean zero and variance $\delta^2 I_{n_i}$, with I_{n_i} as identity matrix, i.e. $\varepsilon_i(t) \sim N\left(0, \delta^2 I_{n_i}\right)$, for all $t \geq 0$.). The model assumes that b_i and $\varepsilon_i(t)$ are independent. Here, b_i accounts for the association between the longitudinal and the event process and the correlation between the repeated measurements in the longitudinal outcome.

1.2 Joint Model

Joint model is a combination of longitudinal and survival submodels. The aim under joint modelling is to associate the true unobserved longitudinal outcome $m_i(t)$ with the hazard of an event. The joint model allows to settle that the longitudinal markers are a function of true unbserved longitudinal value $m_i(t)$ and some error (Hevia, 2014). In general form, with longitudinal and survival submodels as shown in Eqs. (1) and (4), the competing risks joint model is given as:

$$\begin{cases} h_{ik}\left((t|\mathcal{M}_i(t)\right) = h_{0k}\,(\mathrm{t})\exp(\,\gamma_k{}^T w_i + \boldsymbol{\alpha_k}\,(m_i(t)) \\ y_i(t) = m_i(t) + \varepsilon_i(t), m_i(t) = \beta X_i^T(t) + Z_i^T(t)b_i \end{cases} \tag{5}$$

In the joint model**,** b_i accounts for association between the longitudinal and failure processes.

1.3 Parameter Estimation in Competing Risks Joint Model

Estimation in joint modelling uses maximum likelihood estimation. The formulation of likelihood function assumes that the survival and longitudinal processes are conditionally independent given the vector of random effects b_i. It also assumes the independence of longitudinal measurements. The likelihood function is given as follows:

$$p\left(T_i, \delta_i | b_i; \theta_t, \beta\right) = \prod_{k=1}^{K} \left[h_{0k}\left(T_i\right) \exp\left(\gamma_k{}^T w_i + \boldsymbol{\alpha_k}\left(m_i(t)\right)\right)\right]^{I(\delta_i = k)} \times \exp\left(-\sum_{k=1}^{K} \int_0^{T_i} h_{0k}(s) \exp\left(\gamma_k{}^T w_i + \boldsymbol{\alpha_k}\left(m_i(s)\right) ds\right)\right) \tag{6}$$

where $p(.)$ is the probability density function, θ_t denotes parameters for event time outcomes and β denotes parameters for fixed effects.The maximazition of the log-likelihood function corresponding to Eq. (6) gives the estimates (Rizopoulos, 2010). Standard numerical integration techniques are used. However, Rizopoulos et al. (2010) use a quasi-network algorithm, a direct maximisation of observed data log-likelihood. The estimation of the random effects is done through Bayesian theorem. As defined by Rizopoulos (2010), with $p(b_i, \theta)$ as prior distribution and $p(T_i, \delta_i | \, b_i; \theta) p(\, y_i(t_{ij}) \mid b_i; \theta)$ as the conditional likelihood, the posterior distribution is given as:

$$p\left(b_i | T_i, \delta_i, y_i; \theta\right) \propto p\Big(\left(T_i, \delta_i | b_i; \theta\right) p\left(\, y_i\left(t_{ij}\right) \, | b_i; \theta\right) p\left(b_i; \theta\right) \tag{7}$$

where $\theta = \left(\theta_t^T, \theta_y^T, \theta_b^T\right)^T$ denotes the parameter vector for the event time outcome, the longitudinal outcomes and the random effects variance-covariance matrix, respectively.

2 Application to Malaria Data

This study emphasises on joint analysis of longitudinal and survival outcomes and compares the joint models and separate models for longitudinal and survival data obtained from randomised controlled trial for malaria interventional study at Chileka Health Centre which is about 19 km from Queen Elizabeth Central Hospital (QECH) Malawi.

2.1 Materials and Methods

The study used secondary data that was collected from a malaria treatment efficacy randomised controlled trial that aimed at evaluating strategies to delay the emergency of resistance to anti-malaria drugs in children by Malawi Liverpool Wellcome Trust and College of Medicine between 2003 and 2006. Children were randomised to four treatment arms: sulfadoxine-pyrimethamine (SP) + placebo, chloroquine (CQ) + sulfadoxine-pyrimethamine, amodiaquine (AQ) + sulfadoxine-pyrimethamine and artesunate (ART) + sulfadoxine-pyrimethamine. They were followed up for a period of 6 weeks. The results for the primary study were published (Bell et al., 2008).

2.2 Study Population and Sample Size

The primary study targeted children aged between 12 and 60 months, weight $\geq$6 g, no feature of severe malaria on enrolment and hemoglobin level 5.0 g/l. The children were recruited at and followed up from Chileka Health Centre which is about 19 km from QECH. The primary study was a double-blinded trial as all members of study team and patients were uninformed of study treatments allocation. The children were assessed on days 0, 7, 14, 28 and 42 and any other day if unwell.

All children provided venous and capillary blood samples on assessment days for parasite microscopy. In the blood samples, biomarkers such as hemoglobin, white cell count, red blood cell count, platelets, creatinine and bilirubin were examined in full blood count. Children were considered pure, with *P. falciparum* parasitaemia parasite density between 2000 and 200,000 parasites per μl. Children were removed from the primary study after enrolment if their full blood count showed severe anaemia, hemoglobin <5 g/dl. On the other hand, during follow-up, withdrawal was based on adverse reactions to the randomised drug, protocol violation and consent withdrawal. The study used data for 101 patients that were part of a primary study to be recruited over a period of 4 years.

2.3 Statistical Analysis Methods

The distribution of categorical variables was summarised using percentages, box-plots and bar plots. Continuous outcomes were summarised using median and means. Longitudinal and survival outcomes were also presented using individual profile plots (spaghetti plots) and cumulative incidence curves, respectively. In multivariate analysis, five models were fitted: two mixed-effects models for parasite count and hemoglobin level, cause-specific hazard model and two competing risks joint models for parasite counts and hemoglobin level.

The comparison of separate models and competing risks joint models was done using log-likelihood estimates and Bayesian information criterion (BIC). A model with large log-likelihood estimate or lower BIC is considered better fit to the data.

The outcome of interest was time to severe malaria/treatment failure. The competing event outcome was time to withdrawal. In the study, withdrawal was due to adverse reaction to the drugs/treatment used, protocol violation and consent withdrawal. The longitudinal outcomes of interest were hemoglobin level (g/dl) and parasite counts. Effects of baseline covariates, age, treatment, gender and body weight were analysed. In the separate cause-specific hazard model, baseline covariates, hemoglobin and parasite counts (collected on day 0) were included in the survival model to learn if there was any association between the longitudinal biomarkers and the survival processes. All data were analysed using R version 3.5.1 and the main package JM. Other packages used were survival, lattice, splines, nlme4, reshape2 and cmprsk. Statistical significance was declared at α-level of 5%.

2.4 Ethical Approval

For the primary study, the study protocol was approved by ethics committees of the College of Medicine Research and Ethics (COMREC), University of Malawi, and Liverpool School of Tropical Medicine. Written informed consent was required from the parent of each child recruited, and the study was explained in parent's preferred language.

3 Results

In this section analysis results for separate and joint models to malaria data from sub-Saharan Africa are presented.

3.1 Descriptive Analysis

Data for 101 patients were available for analysis. There were more male children (57.4%) than female children (42.6%). The mean age of the children was 2.22 years (std = 1.12) with mean body weight of 11.03 kilograms. The median follow-up time was 28 days. Different survival outcomes observed were severe malaria (5.0%), withdrawal (10.9%) and censored (84.1%). The mean parasite count was 6 parasites per microlitre, and 9.38 g/dl hemoglobin level was for data collected on day 0.

The median parasite counts were the same across subjects with different event types. However, the median hemoglobin level was slightly higher for subjects that were censored (Fig. 1).

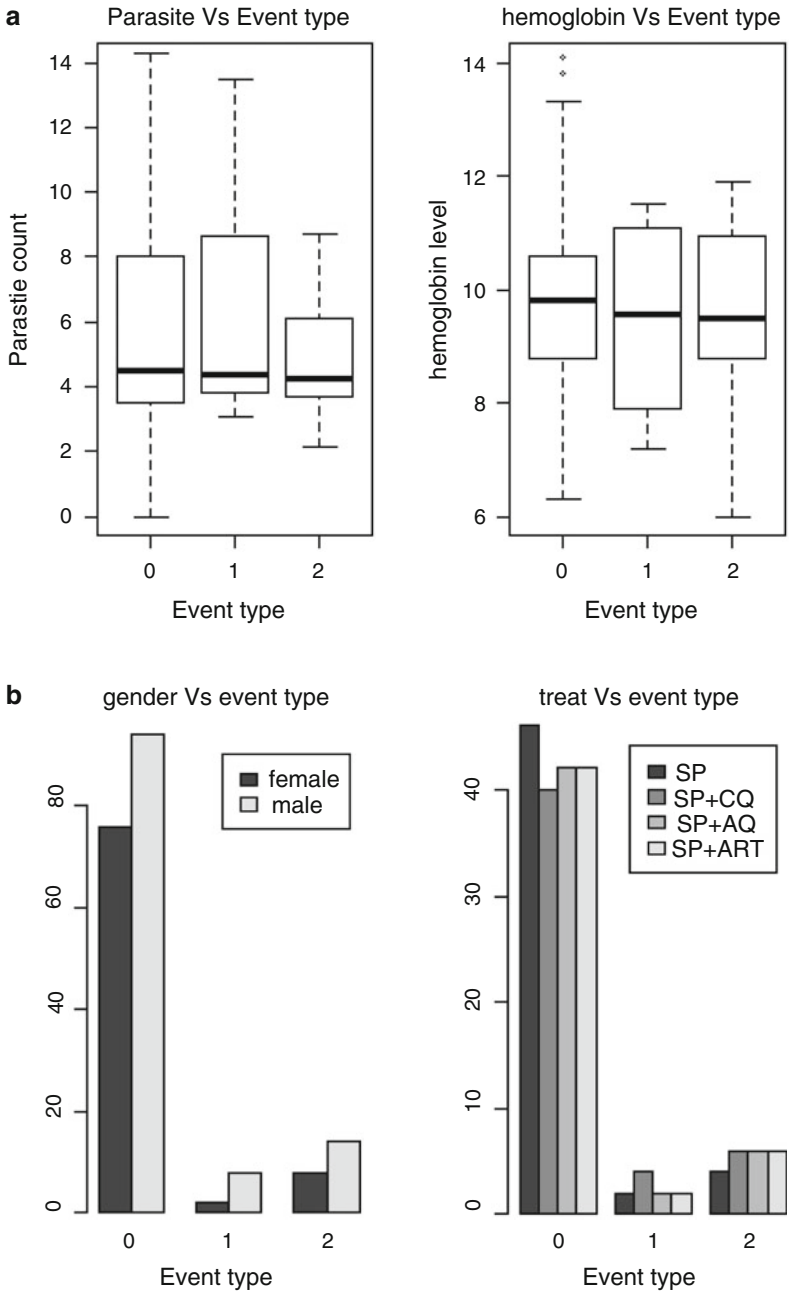


Fig. 1 Baseline variables: (**a**) Parasite count and hemoglobin level against event type. (**b**) Gender and treatment options against event type. Event type (0: censored, 1: severe malaria and 2: withdrawal)

3.2 *Longitudinal Process Analysis*

In the first part, we present subject-specific evolutions in time of the longitudinal biomarkers hemoglobin level and parasite measurements and correlation between parasite count and hemoglobin level. From Fig. 2a, b, it was observed that subjects showed similar variability in their longitudinal profiles for hemoglobin and parasite across treatment groups. The results in Fig. 3 indicate weak positive correlation between the hemoglobin level and parasite count.

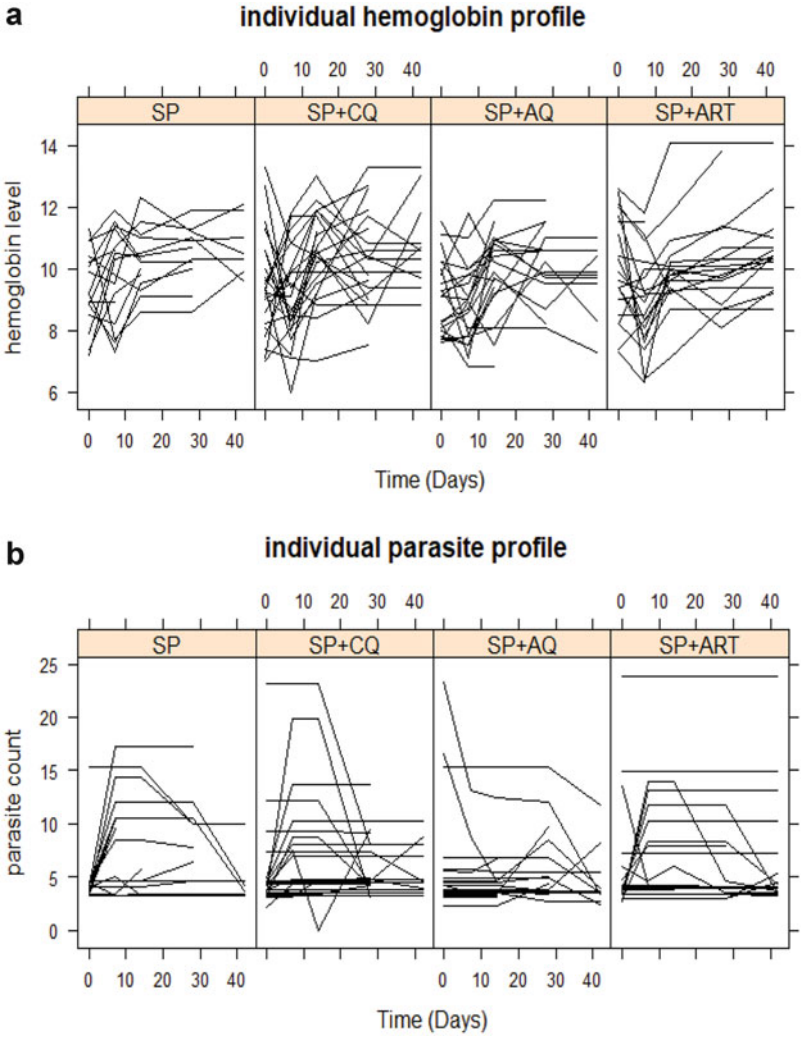


Fig. 2 (**a**) The Individual hemoglobin profile over time in days separated by treatment options. (**b**) The individual parasite count profiles over time in days separated by treatment options

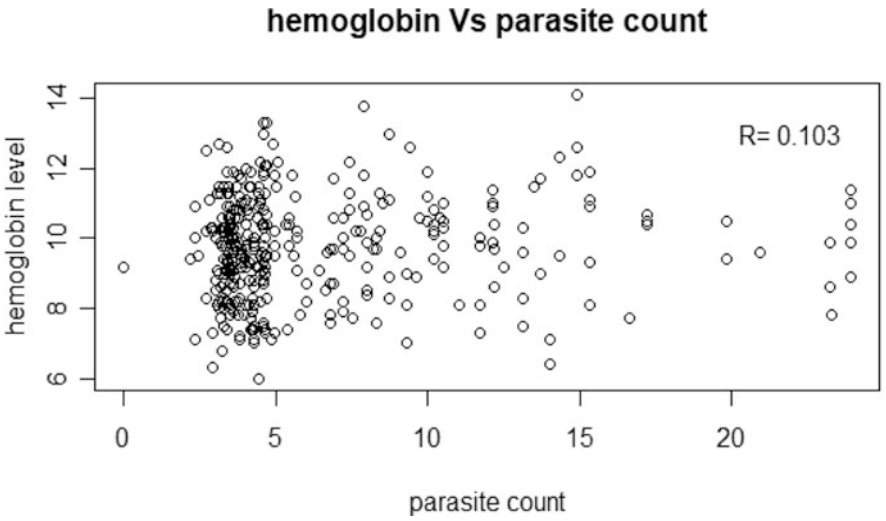


Fig. 3 Scatter plot showing correlation between hemoglobin level and parasite count

Next are results from a fitted mixed-effects models of hemoglobin level and parasite counts presented in Table 1. The results showed that time in days was statistically significant in predicting the longitudinal scores of hemoglobin levels which increased by 0.03 gram/dl for each passing day ($p < 0.0001$). For the parasite count scores, the intercept significantly increased the parasite counts by 5.61counts in the absence of the covariates ($p = 0.016$). The other covariates, treatment, sex, age and body weight, had no significant effects on both hemoglobin level and parasite counts. Looking at behaviour of parasite count profile in Fig. 2b, a quadratic mixed-effects model in terms of time evolution for parasite count was also fitted. The quadratic model showed statistical significance of observed time and square of observed time and the intercept with p-values 0.002, <0.0001 and 0.023, respectively. However, when the two models were compared using BIC, the linear model (BIC = 2035) was opted as the best fit to the data over the quadratic model (BIC = 2036) (Tables 2 and 3).

3.3 Survival Process Analysis

For the survival process, we first looked at cumulative incidence curves for the two competing events, severe malaria and withdrawal. The results showed that the cumulative incidence rates for withdrawal were slightly higher than severe malaria as shown in Fig. 4. When we compared overall survival of female and male patients, the results showed that female patients had slightly higher survival rate beyond day 10 of follow-up as shown in Fig. 5.

Table 1 Fitted values for the linear mixed-effects models for the longitudinal variables hemoglobin level and parasite counts with standard deviations (se) and the *p*-values

Hemoglobin level			
Covariate	Slope	SE	*p*-value
$\boldsymbol{\beta_{H,0}}$ (intercept)	8.38	0.65	<0.0001
$\boldsymbol{\beta_{H,1}}$ (sexmale)	−0.05	0.22	0.835
$\boldsymbol{\beta_{H,2}}$ (age)	0.12	0.14	0.376
$\boldsymbol{\beta_{H,3}}$ (weight)	0.06	0.07	0.409
$\boldsymbol{\beta_{H,4}}$ (time)	0.03	0.004	<0.0001
$\boldsymbol{\beta_{H,5}}$ (SP + CQ)	0.19	0.31	0.549
$\boldsymbol{\beta_{H,5}}$ (SP + AQ)	−0.13	0.32	0.695
$\boldsymbol{\beta_{H,5}}$ (SP + ART)	0.20	0.32	0.528
Random effects			
$\mathbf{b_{H,0i}}$ (intercept)		0.89	
Residuals		1.10	
Parasite count			
$\boldsymbol{\beta_{p,0}}$ (intercept)	5.61	2.30	0.016
$\boldsymbol{\beta_{p,1}}$ (sexmale)	1.55	0.79	0.054
$\boldsymbol{\beta_{p,2}}$ (age)	−0.46	0.49	0.345
$\boldsymbol{\beta_{p,3}}$ (weight)	0.08	0.24	0.744
$\boldsymbol{\beta_{p,4}}$ (time)	−0.01	0.01	0.493
$\boldsymbol{\beta_{H,5}}$ (SP + CQ)	0.30	1.10	0.786
$\boldsymbol{\beta_{H,5}}$ (SP + AQ	−0.59	1.14	0.608
$\boldsymbol{\beta_{H,5}}$ (SP + ART)	0.65	1.13	0.568
Random effects			
$\mathbf{b_{p,0i}}$ (intercept)		3.52	
Residuals		2.76	

Table 2 Fitted quadratic mixed-effect model for parasite count

Parasite count (quadratic)			
Covariate	Slope	SE	*p*-value
$\boldsymbol{\beta_{p2,0}}$ (intercept)	5.62	2.30	0.023
$\boldsymbol{\beta_{p2,1}}$ (sexmale)	1.50	0.79	0.063
$\boldsymbol{\beta_{p2,2}}$ (age)	−0.45	0.45	0.350
$\boldsymbol{\beta_{p2,3}}$ (weight)	0.07	0.24	0.770
$\boldsymbol{\beta_{p2,4}}$ (time)	0.07	0.04	0.002
$\boldsymbol{\beta_{p2,4_*}}$ (time-squared)	−0.04	0.01	<0.0001
$\boldsymbol{\beta_{p2,5}}$ (SP + CQ)	0.10	1.10	0.930
$\boldsymbol{\beta_{p2,5}}$ (SP + AQ)	0.73	1.14	0.524
$\boldsymbol{\beta_{p2,5}}$ (SP + ART)	0.58	1.13	0.609
Random effects			
$\mathbf{b_{p2,0i}}$ (intercept)		3.54	
Residuals		2.68	

Table 3 Fitted cause-specific hazard model for time-dependent variables (Hemoglobin and Parasite count)

Parameter	RR	SE	*p*-value
CQ + SP	1.30	1.27	0.841
CQ + SP: CR	0.91	1.57	0.954
AQ + SP	0.42	1.50	0.560
AQ + SP: CR	3.61	1.77	0.468
ART+ SP	0.41	1.57	0.574
ART+ SP:CR	3.11	1.84	0.538
Age	0.29	0.85	0.153
Age: CR	3.12	0.93	0.223
Weight	1.26	0.36	0.522
Weight: CR	0.99	0.41	0.995
Sexmale	2.90	1.21	0.380
Sexmale: CR	0.48	1.37	0.589
Hb0	0.90	0.29	0.721
Hb0: CR	1.33	0.36	0.432
Parasite0	1.08	0.07	0.259
Parasite0: CR	0.93	0.09	0.477

CR: competing risk

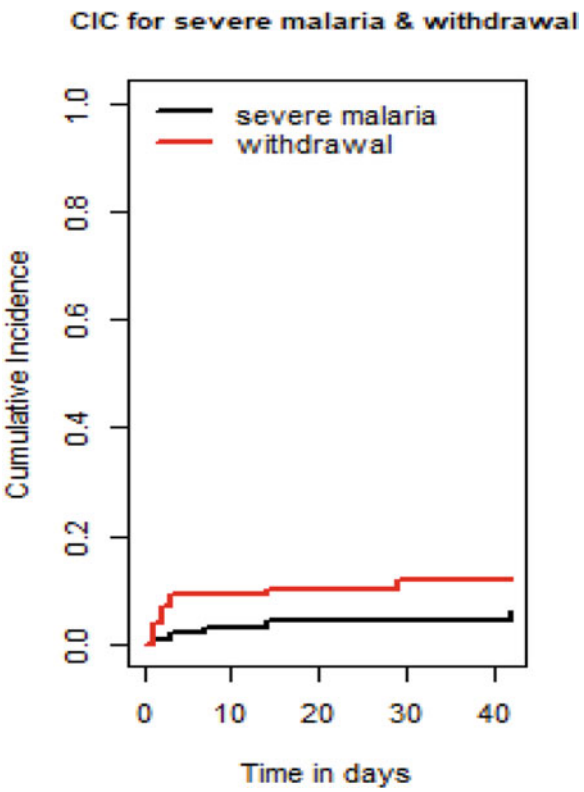


Fig. 4 Cumulative incidence curves for the two competing events, severe malaria and withdrawal

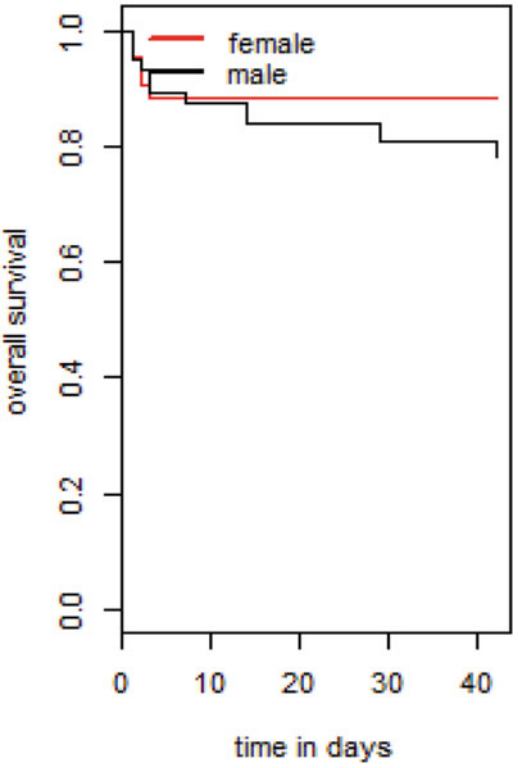


Fig. 5 Kaplan-Meier plots of survival probabilities for male and female patients

In order to determine effects of baseline covariates and the two biomarkers parasite count and hemoglobin level on the risks of severe malaria and withdrawal, a cause-specific hazard model for time-dependent covariates was fitted. The results showed that the baseline hemoglobin level and parasite counts were not statistically significant in predicting the risks of severe malaria and withdrawal in the cause-specific time-dependent hazard model. The covariates, age, body weight, treatment and sex were not statistically significant in predicting the hazards for the two competing events.

3.4 Competing Risks Joint Models for Parasite Count and Hemoglobin Level

Two competing risks joint models were fitted. The first joint model combined the survival and longitudinal processes for true hemoglobin level to model its effect on the risks of severe malaria and withdrawal. As shown in Eq. (5), each joint model is a composition of longitudinal and event/survival processes. The results

Table 4 Estimates for a fitted joint model for longitudinal marker hemoglobin and competing risks survival processes

Parameter	RR	SE	*p*-value
Event process			
CQ + SP	1.15	1.19	0.9043
CQ + SP: CR	1.50	1.64	0.8062
AQ + SP	0.39	1.46	0.5742
AQ + SP: CR	3.93	1.88	0.4659
ART+ SP	0.42	1.42	0.5430
ART+ SP:CR	5.83	1.83	0.3354
Age	0.37	1.82	0.2232
Age: CR	6.62	0.87	0.0305
Weight	1.13	0.34	0.7258
Weight: CR	0.94	0.38	0.8782
Sexmale	3.20	1.16	0.3175
Sexmale: CR	0.65	1.35	0.2927
Assoct: (true hemoglobin level)	0.44	1.04	0.4239
Assoct: CR	0.68	0.73	0.0111
Longitudinal process (hemoglobin level)			
Slope			
Intercept	8.42	0.580	<0.0001
Time	0.03	0.004	<0.0001
Sexmale	−0.05	0.202	0.8015
Age	0.13	0.120	0.2994
Weight	0.05	0.059	0.3745
CQ + SP	0.13	0.282	0.6347
AQ + SP	−0.21	0.290	0.4634
ART+ SP	0.14	0.290	0.6222

showed that for each passing day, there was a unit increase in hemoglobin level by 0.03 g/dl ($p < 0.0001$). The covariates, age, sex, treatment and body weight were not statistically significant in predicting the hemoglobin level scores in the patients. However, in the survival process, results showed that age was significantly associated with the risk of withdrawal (medical grounds) with older children with 13% more likely to withdrawal than younger children (p-value $= 0.0305$). The results also showed that the true hemoglobin level was significantly associated with the risk of withdrawal such that the risk of withdrawing from the study was 22% lower for children with higher hemoglobin level than children with lower hemoglobin level (p-value $= 0.0111$). All the baseline covariates and true hemoglobin level were not associated with the risk of severe malaria in the joint model. The results are shown in Table 4.

The other fitted joint model was for parasite count as a biomarker. In the model, the longitudinal process showed that sex was statistically significant in prediction of parasite count with 1.57 more counts in male patients than female patients

Table 5 Estimates for fitted joint model for parasite count

Parameter	RR	SE	*p*-value
Event process			
CQ + SP	1.22	1.22	0.870
CQ + SP: CR	1.19	1.55	0.909
AQ + SP	0.53	1.45	0.660
AQ + SP: CR	3.42	1.76	0.484
ART+ SP	0.48	1.45	0.611
ART+ SP:CR	4.08	1.75	0.421
Age	0.37	0.74	0.175
Age: CR	2.45	0.83	0.282
Weight	1.11	0.32	0.755
Weight: CR	1.15	0.37	0.706
Sexmale	3.74	1.17	0.259
Sexmale: CR	0.36	1.34	0.449
Assoct: true parasite	0.92	0.17	0.632
Assoct: CR	1.05	0.21	0.813
Longitudinal process (parasite count)			
Slope			
Intercept	5.61	2.23	0.012
Time	−0.008	0.23	0.470
Sexmale	1.57	0.77	0.041
Age	−0.46	0.47	0.326
Weight	0.08	0.23	0.741
CQ + SP	0.32	1.07	0.765
AQ + SP	−0.61	1.10	0.582
ART+ SP	0.65	1.09	0.556

Table 6 Estimates for Bayesian information criteria and log-likelihood for the models

Model	BIC	Log-likelihood
Separate HB level	1306.7	−623.91
Joint HB level	1583.60	−704.11
Separate parasite count	2035.4	−988.24
Joint parasite count	2316.66	−1070.64
Cause-specific model	176.93	−66.29

(p-value = 0.041) which can be better explained by increased overall survival in female patients than male patients as shown in Fig. 4. The survival process showed that the baseline covariates age, gender, weight and treatment were not statistically significant in predicting the risks of withdrawal and severe malaria in the joint model. The true parasite count was not significantly associated with the two competing risks severe malaria and withdrawal. The results are shown in Table 5.

3.5 Model Comparison

As the data were fitted using separate models for analysis of longitudinal and survival data and joint modelling framework, the models were compared to determine which models better fitted the malaria data. In the analysis, Bayesian information criteria was used to compare the models. The results indicated that due to lack of association between the longitudinal and survival processes, it would be enough to analyse the data using separate models for longitudinal and survival data analysis. In all cases, the BIC for separate models was lower compared to BIC for joint models. The same conclusions were reached when log-likelihood was used for comparison of the models. The results are shown in Table 6.

3.6 Model Diagnostics

In joint models alone, Rizopoulos (2012) recommended the use of multiple imputation residuals with the fixed visit times to validate the model assumptions. However, when the survival, longitudinal and competing risks components are amalgamated, model assessment is complicated, and methods were not available in JM package during analysis. As this work includes competing risks, model diagnostics in competing risks joint models were not done.

4 Discussion

This study has found that time was significant in predicting hemoglobin level. This is a common effect that one would expect in biomedical research for malaria. As time passes, the hemoglobin level increases under normal conditions. This could be the case as patients kept on getting malaria medication that in turn reduces the parasite count. However, this finding cannot be generalised as with passing days; existence of other conditions may affect the hemoglobin level (White, 2017). The positive effect of time is also shown in competing risks joint model. As the survival process for hemoglobin level suggested, age was associated with the risk of withdrawal with older children more likely to withdrawal (medical condition) from the study than younger children. This result is not well presented biologically as withdrawal included guardian's decisions when they felt the child is unwell. The results showed that true hemoglobin level was also associated with the risk of withdrawal with higher hemoglobin level children withdrawing (medical grounds) than lower hemoglobin level children. This could be the case, as children with higher hemoglobin level could signal recovery from malaria, hence having their parents denying consent to proceed with the study.

For the parasite count models, it is only in the joint modelling framework where sex of the child is significantly associated with parasite count scores. Male patients had more parasite count than female children. As indicated by analysis, female children had an increased overall survival of severe malaria than male children. A study by Kotepui et al. (2014) found significant differences in malaria cases between male and female people.

In the analysis of these data, comparison of separate hemoglobin longitudinal model and the joint model longitudinal process revealed that in both models time was significant having the same effect in joint and separate models. However, reduced standard deviations were observed in joint model than separate model suggesting biased estimate in separate model (Rizopoulos, 2012). For the parasite count models, no covariate was significant in predicting the longitudinal scores of parasite counts in the mixed-effects model. The case was different in the joint model, where sex was significant with smaller standard error for the estimate in joint model. This could be the result of including the true longitudinal parasite counts in the joint process, leading to some covariates having an effect on parasite counts.

In the joint model for the parasite counts and competing risks, the covariates weight, sex and true parasite count had no significant effects on the risks of severe malaria and withdrawal. However, the estimates had smaller standard errors for joint model estimates, than the separate cause-specific model and mixed-effect model. The large standard errors in separate model could disadvantage the separate models since in practice models with small standard errors receive more attention (Nguti et al., 2005).

The results showed that separate models were more applicable to these data than joint models. The preference of separate models in this work to joint models is indicated by lack of association between the longitudinal biomarkers and the survival outcomes. However, in literature, joint models considering competing risks by Hevia (2014); Hickey, Philipson and Jorgensen (2018); and Andrinopoulou (2014) were preferred for analysis of data. The choices could be understood as there were associations between the longitudinal markers of interest in each study and the event or survival processes in separate time-dependent Cox models. In the work by Hevia (2014), the joint model indicated smaller standard errors than the separate model, hence giving the joint models more preference.

The goals for this work were to introduce joint modelling framework as it gains popularity and wider applications in biomedical research and apply the models to randomised controlled trials for malaria study. The work was done with data collected many years ago with a small sample size, and there may be possible trends in the current data for malaria studies. We recommend use of latest large malaria datasets for updates and appreciate the performance of joint models. We also recommend simulation of data at different correlation levels between the survival and longitudinal data to validate the performance of joint models whenever the two processes are interrelated.

We observed that in biomedical studies, where statistical tools are used, further progress is needed in this area of joint modelling of longitudinal data with competing risks survival data to advance tools for better analysis, as the field is in

early developmental stages and restricted in its application to biomedical studies. For instance, there is a need for development of diagnostic methods for model validation in competing risks joint modelling settings, selection and comparison. Multivariate competing risks joint models are another area for further study as longitudinal outcomes coexist.

5 Conclusion

When analysing the longitudinal malaria outcomes together with competing risks survival outcomes in randomised controlled trials, joint models can be considered where there is an association between the longitudinal and events processes. For these malaria data, there was no association between any of the longitudinal biomarkers and the risk of severe malaria and withdrawal in the extended time-dependent cause-specific hazard models. Therefore the separate models could be preferred to analyse these malaria outcomes data than the joint models with competing risks as association may not be of interest. However, correlated data for longitudinal and survival processes could give a different choice of the model preference.

Acknowledgements We thank SACCAB DELTAS Africa for their financial support. We also thank Dr. David Bell for allowing us to use the data.

Competing Interests The authors declare that they have no competing interests.

Author's Contribution	DM	AM	MM
Research concept and design	✓	✓	✓
Collection of data	✓	✓	✓
Data analysis and interpretation	✓	✓	✓
Writing the article	✓	✓	✓
Critical revision of the article	✓	✓	✓
Statistical analysis	✓	✓	✓
Final approval of the article	✓	✓	✓

References

Andrinopoulou, E.-R. (2014). *Joint modelling of longitudinal and survival data with applications to heart valve data*. ISBN, 978–994.

Bell, D. J., Nyirongo, S. K., Mukaka, M., Zijlstra, E. E., Plowe, C. V., Molyneux, M. E., & Winstanley, P. A. (2008). Sulfadoxine-pyrimethamine–based combinations for malaria: A randomized blinded trial to compare efficacy, safety and selection resistance in Malawi. *PLos One, 3*(2), e1578. https://doi.org/10.1371/journal.pone.0001578

Boor, C. d. (1978). *A practical Guide to splines*. Springer.
Collet, D. (2003). *Modelling Survival data in Medical Research*. Chapman & Hall/CRC.
Elashoff, R. M., Li, G., & Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics, 64*, 762–771.
Faucett, C., & Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine, 15*, 1663–1685.
Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. A John Wiley & Sons.
Hevia, A.F. (2014) Joint modeling of longitudinal and survival data.. An application to biomedical research available on: http://www.eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto.
Hickey, G., Philipson, P., Jorgensen, A., & Kolamunnage-Donà, R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 181*. https://doi.org/10.1111/rssa.12348
Hsieh, F., Tseng, Y.-K., & Wang, J.-L. (2006). Joint modelling of survival and longitudinal data: Likelihood approach revisited. *Journal of the International Biometic Society, 62*, 1037–1043.
Ibrahim, J. G., Chu, H., & Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Clinical Oncology, 28*, 2796–2801.
Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text*. Springer.
Kotepui, M., Phunphech, B., Phiwklam, N., Chupeerach, C., & Dvangmano, S. (2014). Effects of Malaria Infection on haematological parameters in population near Thailand, Myanmar border. *Malaria Journal, 13*, 218.
Molenberghs, G., & Kenward, M. G. (2007). *Missing data in clinical studies*. Wiley.
Nguti, R., Burzykowski, T., Rowlands, J., Renard, D., & Janssen, P. (2005). Joint modelling of repeated measurements and event-time: An application to performance traits and survival of lambs bred in sub-humid tropic. *Genetic Selection Evolution: Biomedical Central, 37*(2), 175–197.
Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software, 35*(9), 1–33.
Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data with applications in R*. CRC Press.
Sudell, M., Kolamunnage-Dona, R., & Tudu-Smith, C. (2016). Joint models for longitudinal and time-to-event data: A review of reporting quality with a view to meta-analysis. *BMC Medical Research Methodology, 16*, 168.
Tsiatis, A., DeGruttola, V., & Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association, 90*, 27–37.
White, N. (2017). Malaria parasite clearance. *Malaria Journal, 16*, 194.
Williamson, P. R., Kolamunnage-Dona, R., Philipson, P., & Marson, A. G. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in Medicine, 27*(30), 6426–6438.
Yu, M., Law, N. J., Taylor, J. M. G., & Sandler, H. M. (2004). Joint longitudinal survival-cure models and their application to prostate cancer. *Statistica Sinica, 14*(3), 835–862.

Stratified Multilevel Modelling of Survival Data: Application to Modelling Regional Differences in Transition to Parenthood in Ethiopia

Gebrenegus Ghilagaber, Amanda Akinyi Lagehäll, and Elelta Yemane

Abstract This chapter presents a multilevel extension of the Cox proportional hazards model where a shared frailty term is included to account for clustering of women within households. The extended model is used to analyze regional differences in the intensity of transition to parenthood among 15019 Ethiopian women aged 15–49 years old in the country's Demographic and Health Survey of 2016. Women's birth cohort, residence and educational level were used as background variables. Conventional Cox proportional hazards models and two multilevel models (with gamma distributed and log-normal distributed frailty terms) are fitted to data for the entire country and, separately, for each of the nine regions and two city administrations. We found that household frailty effects are fairly small in the nine regions but the log-normal frailties were significant in the entire country and the two city administrations which are relatively heterogeneous with inhabitants from many ethnic groups. We also found regional differences in the effects of the background variables on the intensity of transition to parenthood but the effects were generally stable across the three models in each region. Overall, we recommend use of multilevel survival models to account for clustering of women into households and proper care in the choice of distribution of the household random effects.

Keywords Unobserved heterogeneity · Right-censored data · Frailty models · Shared frailty · R · Survival package in R · Correlated survival data · Hazard/intensity functions · Proportional hazards models · Partial likelihood · Penalized likelihood · Baseline hazards · Semi-parametric inference · Gamma distribution · Log-normal distribution · Ethiopia · Transition to parenthood · Fertility · Demographic and Health Surveys (DHS) · Education · Residence · Birth cohort · Regional differences · Time-to-event data · Intra-class correlation coefficient · Stratified modelling · Timing of first birth · Clustered data · Multilevel modelling · Tigray region · Afar region · Amhara region · Oromia

G. Ghilagaber (✉) · A. Akinyi Lagehäll · E. Yemane
Department of Statistics, Stockholm University, Stockholm, Sweden
e-mail: Gebre@stat.su.com

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_17

region · Somali region · Benishangul-gumuz region · Southern nations, Nationalities and peoples region · Gambela region · Harari region · Addis abeba city administration · Dire dawa city administration

1 Introduction

Individuals are often clustered within groups, such as children within mothers, students within classes and schools, and schools within geographical areas or regions. This is the case in Demographic and Health Surveys where, for instance, women are clustered within households. Since women in the same household are expected to be more alike than women selected at random from the population, the basic assumption of a random (independent) sample of women cannot be ascertained. Further if, as would be expected, women within the same household are positively correlated then ignoring the clustering may lead to overestimation of the precision of estimates of covariates in subsequent analyses. Thus, analytical methods need to pay due attention to the hierarchical nature of the data. This can be achieved by using the clusters (households) as units of analysis instead of the individuals (women) and treating women from the same household as correlated cases (multi-levels) within the same observation (household) and use some form of multilevel modelling that are now abundant in the literature. Such procedure has the advantage of accounting for any household-specific unobserved heterogeneity that may affect outcome at a woman level.

In this chapter we address these issues in the context of regional differences in the intensity of transition to parenthood among Ethiopian women based on data from the country's Demographic and Health Survey (DHS) of 2016. The survey questionnaire included a complete birth history, as well as information on some background variables like women's education, residence, birth cohort, and related covariates. The resulting usable records of 15019 women were clustered into 4722 households, the number of women in a household varying between 1 and 15 while the mean number of women varying between 2.27 and 4.06 woman per household.

We analyze the above clustered data using a multivariate survival models that allow for correlated observations. We begin with the standard Cox proportional intensities model that ignores clustering in estimating the effects of observed covariates (residence, birth cohort, and education) on intensity of transition to parenthood. We then add a household-specific random effect that represents unobserved influences common to all women in a household.

Such models and their variants are already available in the literature. For instance, Vaupel et al. (1979) have used random effects to represent the effects of unobserved population heterogeneity in survival models. Clayton (1978) proposed a bivariate intensity model that can be interpreted in terms of a proportional intensities model with a gamma distributed random effect while Heckman and Singer (1982, 1984) and Trussell and Richards (1985) have studied the sensitivity of parameter estimates to the choice of distribution for the random term. Guo and Rodriguez (1992), Sastry

(1997), and Bolstad and Manda (2001) investigate family and community random effects on child survival in Guatemala, Northern Brazil, and Malawi, respectively. McGilchrist and Aisbett (1991); McGilchrist (1993) propose estimation methods and applications in situation of repeated events.

The focus of this chapter is to examine, in the framework of shared frailty model, if there are household random effects on the intensity of transition among Ethiopian women and if such random effects differ across the different regions in the country. Further, we also investigate if the effects of background covariates on transition to parenthood are sensitive to the choice of distribution for the random effect.

The rest of the chapter is organized as follows. In Sect. 2, we introduce the data set and its structure. Section 3 describes proportional intensities models. Both the standard Cox proportional intensities model and two extensions (with gamma distributed and log-normal distributed random effects, respectively) are presented. In Sect. 4, we present and discuss empirical results from fitting the models of Sect. 3 to the data of Sect. 2 and provide summary and concluding remarks in the last section.

2 The Data Set: Distribution Across Regions and Covariates and Clustering Within Households

2.1 Ethiopia and Its Regions

Ethiopia is located in East Africa and is bordered by South Sudan and Sudan to the west, Djibouti to the east, Eritrea to the north, and Kenya and Somalia to the south. Since the early 1990s the country is partitioned into 9 ethnic/linguistic regions and two city administrations. In the north of the country we have the Tigray, Afar, and Amhara regions while Benishangul-Gumuz and Gambela are in the west of the country. The Southern Nations Nationalities and Peoples Region (SNNPR) forms the southern part of the country. A new region (Sidama region) has been formed as a split from the SNNPR in June 2020. To the east of the country we have the Ethiopian Somali region and the Harari region. Oromia is a large region stretching from the south to the central and the western part of the country. The two city administrations are the capital city, Addis Abeba, which is in the center of the country and the Dire Dawa city administration which is adjacent to the Harari region in the east.

According to the 2016 Ethiopian Demographic and Health Survey (Central-Statistical-Agency-Ethiopia and ICF, 2016), the percentage of young women aged 15–19 who have begun childbearing varied between 3% in the capital city Addis Abeba and 23% in the Afar region located to the northeast. The overall fertility rate also varied between different regions from a total fertility rate (TFR) of 1.8 in the capital city, Addis Abeba, to 7.2 in the Somali region in the eastern part of the country.

2.2 The 2016 Ethiopia Demographic and Health Survey

The data analyzed in this chapter come from the 2016 Ethiopia Demographic and Health Survey (Central-Statistical-Agency-Ethiopia and ICF, 2016). The survey was conducted between January and June 2016 and interviewed 15683 women aged 15 to 49 years. These were considered to be nationally representative sample to provide estimates at national and regional levels as well as for urban and rural areas. The sample frame used in the collection of data was the Ethiopian Population and Housing Census which is a complete list of 84,915 enumeration areas created in 2007. An enumeration area is a geographic area containing 181 households on average. The sample frame contains information on the type of residence area, location, and the estimated number of residential households. The usable records for the purpose of this chapter are 15019 women with available information on the variables of interest.

2.3 Response and Explanatory Variables

The response variable in this study is the rate at which women get their first child (make transition to parenthood). Women are at "risk" of transition once they turn 15 years (the few women who reported having their first birth before they turned 15 years were deleted). Exposure time is the number of months from age 15 until first birth or the survey data, whichever comes first. A woman's survival time represents event time (years between age 15 and first birth) if she was parent by the survey time or censoring time (years between age 15 and the survey date) if she has not yet got her first child at the time of the survey.

Woman's educational level, her residence, and birth cohort have been included as explanatory variables (covariates). Even region has been used as a covariate while analyzing the data for the entire country.

Distribution of the 15,019 women across regions and the three covariates is presented in Table 1. Thus, 1612 of the 15,019 women (10.73%) were from the Tigray region, 1084 (7.22%) were from the Afar region, 1793 (11.94%) were from the Oromia region, and the same number of women were from the capital city, Addis Abeba, etc.

At the bottom of Table 1 we provide the number of households into which the women from each region are clustered in while Table 2 shows frequency distribution of households by number of women (size of household). The last column in the lower panel of Table 2 shows that 1017 of the 4722 households in the sample (21.54%) contributed one woman each, 1090 households (23.08%) contributed two women each, 926 households (19.61%) contributed three women each, etc.

The percentage of transition to parenthood varied across the regions, ranging from 41% in the capital Addis Abeba to 73% in the Afar region. The number of women per household also varied between the regions and across the covariates.

Table 1 Distribution of women across covariates and regions: Ethiopia DHS 2016

Covariate	Ethiopia	Tigray	Afar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis Abeba	Dire Dawa
Urban	5166	399	186	235	243	323	135	219	322	501	1793	810
Rural	9853	1213	898	1399	1550	1024	911	1569	636	369	–	284
Born 1997–2001	3474	423	262	354	411	316	236	391	205	181	431	264
Born 1992–1996	2824	311	227	272	313	250	197	315	186	194	359	200
Born 1987–1991	2714	237	222	298	316	227	185	361	179	148	319	222
Born 1982–1986	2099	180	137	240	283	179	136	256	150	119	263	156
Born 1977–1981	1785	201	97	205	222	163	149	213	111	111	203	110
Born 1972–1976	1201	131	68	143	134	129	90	152	70	76	131	77
Born 1967–1971	922	129	71	122	114	83	53	100	57	41	87	65
No Education	6598	669	790	882	892	1000	517	770	296	299	144	339
Primary Education	5040	524	235	483	692	240	374	765	369	301	649	408
Secondary Education	2197	317	37	189	151	74	98	184	204	156	556	231
Higher Education	1184	102	22	80	58	33	57	69	89	114	444	116
Number of women	15019	1612	1084	1634	1793	1347	1046	1788	958	870	1793	1094
Percent in sample	100	10.73	7.22	10.88	11.94	8.97	6.96	11.91	6.38	5.79	11.94	7.28
Number of households	4722	452	410	452	466	451	415	459	390	383	442	402
Av. women per HH	3.18	3.57	2.64	3.62	3.85	2.99	2.52	3.90	2.46	2.27	4.06	2.72

Table 2 Size of households by region in the 2016 Ethiopian DHS

	Tigray		Afar		Amhara		Oromia		Somali		Benishang.	
Size	HH	Women	HH	Women	HH	Women	HH	Women	HH	Women	*H H*	Women
1	66	66	120	120	61	61	49	49	94	94	126	126
2	90	180	116	232	87	174	96	192	103	206	107	214
3	99	297	70	210	91	273	91	273	109	327	93	279
4	75	300	50	200	90	360	67	268	65	260	48	192
5	49	245	21	105	54	270	72	360	39	195	26	130
6	29	174	21	126	26	156	43	258	30	180	6	36
7	21	147	6	42	24	168	19	133	6	42	7	49
8	12	96	5	40	9	72	15	120	3	24	0	0
9	6	54	1	9	5	45	6	54	1	9	0	0
10	3	30	0	0	2	20	5	50	1	10	2	20
11	1	11	0	0	2	22	2	22	0	0	0	0
12	1	12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	1	13	0	0	0	0	0	0
14	0	0	0	0	0	0	1	14	0	0	0	0
Total	**452**	**1612**	**410**	**1084**	**452**	**1634**	**466**	**1793**	**451**	**1347**	**415**	**1046**
	SNNPR		Gambela		Harari		Addis Abeba		Dire Dawa		Ethiopia	
Size	HH	Women	HH	Women	HH	Women	HH	Women	HH	Women	HH	Women
1	60	60	121	121	128	128	63	63	129	129	1017	1017
2	82	164	114	228	121	242	80	160	94	188	1090	2180
3	83	249	83	249	73	219	63	189	71	213	926	2778
4	71	284	34	136	37	148	82	328	46	184	665	2660
5	55	275	21	105	16	80	45	225	28	140	426	2130
6	49	294	8	48	7	42	46	276	16	96	281	1686
7	29	203	5	35	0	0	22	154	11	77	150	1050
8	19	152	3	24	0	0	18	144	2	16	86	688
9	5	45	0	0	0	0	6	54	1	9	31	279
10	4	40	0	0	0	0	5	50	2	20	24	240
11	2	22	0	0	1	11	4	44	2	22	14	154
12	0	0	1	12	0	0	4	48	0	0	6	72
13	0	0	0	0	0	0	1	13	0	0	2	26
14	0	0	0	0	0	0	0	0	0	0	1	14
15	0	0	0	0	0	0	3	45	0	0	3	45
Total	**457**	**1766**	**390**	**958**	**383**	**870**	**442**	**1793**	**402**	**1094**	**4722**	**15,019**

Thus, a deeper exploration of these regional differences across the covariates and the role of household random effects in these differences is a worthwhile effort. This is what we do in Sect. 4 after introducing proportional intensities model and its shared frailty extensions in the next section.

3 Proportional Intensity Models for Transition to Parenthood

A central concept in the analysis of survival is the hazard or intensity function. Such a function, commonly denoted by $h(t)$ or $\lambda(t)$, is defined as the instantaneous rate at which the event of interest occurs at a specific point t of a (non-negative) time variable T given that it has not occurred before:

$$\lambda(t) = \lim_{\Delta t \longrightarrow 0} \frac{P\left[t < T \leqslant t + \Delta t | T \geq t\right]}{\Delta t}. \tag{1}$$

Intensity functions vary not only over time but also among individuals with different socio-demographic characteristics. Thus, one objective in the analysis of survival data is to draw inferences about the influence of such characteristics (covariates) on the intensity function.

3.1 The Standard Cox Proportional Intensities Model

In his influential paper, Cox (1972) proposed a model where a vector of covariates $\mathbf{x}$ affects the intensity function in a multiplicative manner according to

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp\left(\mathbf{x}'\beta\right), \tag{2}$$

where $\lambda_0(t)$ is an unspecified (non-parametric) baseline intensity function of time and β is an unknown vector of parameters representing the effects of covariate vector $\mathbf{x}$.

Thus, in proportional intensity models (2), the covariates act multiplicatively on the baseline intensity so that their effect is to increase or decrease the intensity of individuals with covariate vector $\mathbf{x}$ relative to the baseline intensity $\lambda_0(t)$ which corresponds to those with $\mathbf{x} = \mathbf{0}$. Further, the ration of intensities of two different individuals with covariates x and x^*, respectively, is constant over time:

$$\frac{\lambda\left(t|\mathbf{x}\right)}{\lambda\left(t|\mathbf{x}^*\right)} = \frac{\lambda_0\left(t\right)\exp\left(\mathbf{x}'\beta\right)}{\lambda_0\left(t\right)\exp\left(\mathbf{x}^{*\prime}\beta\right)} = \exp\left(\left(\mathbf{x}' - \mathbf{x}^{*\prime}\right)\beta\right). \tag{3}$$

This, in turn, implies that the intensity functions of the two different individuals curves are assumed to be parallel over time (cannot cross each other over the study period), something that should and can be tested in any application.

Since no distribution is specified for $\lambda_0\left(t\right)$, one cannot use the usual maximum likelihood approach to estimate the parameter β in (2). Instead, Cox (1975) proposed a partial likelihood approach for estimating β.

Suppose we have pairs of observations on survival times and censoring indicators $\left(t_j, \delta_j\right)$ as well as (fixed) covariates $x_1, \ldots, x_n$, $j = 1, \ldots, n$. No ties are assumed (all t_j are distinct) and an index set of event times is defined for $E \subseteq \{1, \ldots, n\}$:

$$E = \{k | \delta_k = 1\}.$$

Further, for any $t \geq 0$ the risk set (the index set of subjects at risk at time t) is defined as

$$R(t) = \{k | t_k \geq t\}.$$

The partial likelihood is then given by

$$L(\beta) = \prod_{k \in E} \frac{\exp\left(x_k' \beta\right)}{\sum_{m \in R(t_k)} \exp\left(z_m' \beta\right)}. \tag{4}$$

Cox (1975) suggested to estimate β by maximizing the partial likelihood $L(\beta)$ in Eq. (4).

Two special features of the partial likelihood above are (i) it does not depend on the baseline intensity function λ_0 and (ii) it does not depend on the actual event times but only on their rank. The usual (full) likelihood is partitioned into the contributions of the rank and order of the event times. Rank of event times refers to who has the shortest time to the event, who has the next shortest, who has the longest, etc., while the order of the event times refers to the actual values of these times (say the number of months or years to experience the event).

The logic behind the partial likelihood in Eq. (4) is that only the part of the full likelihood based on the rank of event times is enough for the purpose of estimation and inference on β. More details on limited information inference in general (and partial likelihood, in particular) can be found in, for instance, chapter 9 of Lancaster (1990).

Cox (1972) proportional hazards model and its associated partial likelihood in Cox (1975) have dominated statistical applications in many areas over the past 50 years.

3.2 Multilevel Cox Proportional Intensities Model: Accounting for Clustering of Women into Households

3.2.1 Multilevel (Frailty) Models

The conventional Cox intensity model (2) does not account for clustering and, hence, assumes the survival times of n individuals, say $T_1, \ldots, T_n$, to be independent. But, ignoring clustering can induce bias by, for instance, underestimating

standard errors of covariates. Therefore, there have been developments in statistical methods and associated software programs to account for clustering of individuals within groups.

Suppose that we have H households and let $T_{i_1}, \ldots, T_{in_i}$ denote random variables representing the survival times in household i, and let x_{ij} denote a vector of covariates associated with the j^{th} member of the i^{th} household (we assume time-invariant covariates here).

To account for clustering into households a random effect term is included to the model in Eq. (2). The aim is to allow correlations between events times of individuals in the same household. The intensity function for the j^{th} individual in the i^{th} household is then given by

$$\lambda_{ij}(t|\mathbf{x_{ij}}) = \lambda_0(t) \exp\left(\mathbf{x'_{ij}}\beta + \zeta_i\right), \tag{5}$$

where, as before, $\lambda_0(t)$ is an arbitrary baseline intensity function, x_{ij} is the vector of (fixed-effect) covariates, β is the vector of regression coefficients, and ζ_i is the random effect for household i. The vector of covariates x_{ij} may constitute women-specific variables or household-specific variables (with the same value for all women in the same household)

Frailties are the exponential transformations of the random components and the frailty model can be written as

$$\begin{aligned} \lambda_{ij}(t) = & \quad \lambda_0(t) \exp\left(\mathbf{x'_{ij}}\beta + \zeta_i\right) \\ = & \quad \exp(\zeta_i)\, \lambda_0(t) \exp\left(\mathbf{x'_{ij}}\beta\right) \\ = & \quad Z_i \lambda_0(t) \exp\left(\mathbf{x'_{ij}}\beta\right), \end{aligned} \tag{6}$$

where $Z_i = e^{\zeta_i} = \exp(\zeta_i)$, $i = 1, \ldots, H$ are the frailties. The random components $\zeta_1, \zeta_2, \ldots, \zeta_H$ (alternatively, the frailties $e^{\zeta_1}, e^{\zeta_2}, \ldots, e^{\zeta_H}$) are assumed to be independent and identically distributed.

In a shared frailty model, frailty is defined as a measure of the relative intensity that women within a household share. In other words, the frailty variable is associated with groups of women rather than with individual women. Thus, women in household i and with the same values on observed covariates tend to experience transition to parenthood at a faster rate (if $Z_i = e^{\zeta i} > 1$) or slower rate (if $0 < Z_i = e^{\zeta i} < 1$) than they would have done under an independence model in Eq. (2) where no account is made for clustering of women into households.

Modelling is based on the random effects rather than on the frailties. Two frailty distributions, the gamma and log-normal distributions, are common in frailty models.

3.2.2 Cox Proportional Hazards Model with Gamma Distributed Frailty

When Z_i is assumed to follow a gamma distribution with mean $E(Z_i) = 1$ and variance $Var(Z_i) = \theta$, $Z_i = e^{\zeta_i} \sim G\left(\frac{1}{\theta}, \frac{1}{\theta}\right)$, then the density function of the random components ζ_i is given by

$$f(\zeta_i, \theta) = \frac{\exp\left(\frac{\zeta_i}{\theta}\right) \exp\left(-\frac{\exp(\zeta_i)}{\theta}\right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)}, \tag{7}$$

or, equivalently, the density of the frailty term Z_i is given by (see Wienke (2011), page 140)

$$f(Z_i, \theta) = \frac{Z_i^{\frac{1}{\theta}-1} \exp\left(-\frac{z_i}{\theta}\right)}{\theta^{\frac{1}{\theta}} \Gamma\left(\frac{1}{\theta}\right)}. \tag{8}$$

Following Wienke (2011) the corresponding joint likelihood is given by

$$\prod_{i=1}^{H} \left[\int_0^{\infty} \left\{ \prod_{j=1}^{n_i} \left[z_i \lambda_0(t_{ij}) e^{\boldsymbol{\beta}' \mathbf{X}_{ij}} \right]^{\delta_{ij}} \exp\left[-z_i \Lambda_0(t_{ij}) e^{\boldsymbol{\beta}' \mathbf{X}_{ij}} \right] f(Z_i, \theta) \right\} dz_i \right], \tag{9}$$

where

$$\Lambda_0(t) = \int_0^t \lambda_0(s) ds \tag{10}$$

denotes the cumulative baseline hazard function, $f(Z_i, \theta)$ is as defined in Eq. (8) and δ_{ij} is a censoring indicator for members of household i, with $\delta_{ij} = 0$ for censored individuals, $\delta_{ij} = 1$ for individuals who experienced the event of interest (transition to parenthood).

The unknown parameter θ is a dispersion parameter (variance) of the frailties. Thus, each frailty distribution has mean 1 and variance θ. According to Guo and Rodriguez (1992), Oakes (1982) has shown that the parameter θ is closely related to Kendall (1962) coefficient of rank correlation τ. Accordingly,

$$\frac{\theta}{2+\theta},$$

in the context of Eq. (5), can be interpreted as a measure of intra-cluster (intra-household) rank correlation after adjusting for observed covariates. Austin (2017) also used this relationship.

If the variance θ is 0, then women within the same household are mutually independent and Eq. (5) reduces to Eq. (2).

Procedures available in software like Therneau and Lumley (2021) survival-package in R estimate the regression coefficients β, the random effects $\zeta_1, \zeta_2, \ldots, \zeta_n$, and the variance θ using a penalized partial likelihood (Therneau et al., 2003).

3.2.3 Cox Proportional Intensities Model with Log-Normal Distributed Frailty

An alternative distribution for the random effect term is the log-normal distribution. When the frailties $Z_i = e^{\zeta_i}$ follow a log-normal distribution then the $\zeta_i = \log(Z_i)$ follow a normal distribution with mean $E(\zeta_i) = 0$ and variance $Var(\zeta_i) = \theta$, $\zeta_i \sim N(0, \theta)$. The corresponding density function of ζ_i is given by

$$f(\zeta_i, \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{\zeta_i^2}{2\theta}\right). \tag{11}$$

Again, the unknown parameter θ is a dispersion parameter (variance). In other words, each frailty distribution has a mean 0 and variance θ.

Our illustration in the next Section is based on the gamma and log-normal frailty models. Other frailty models in the literature include the shared positive stable frailty model and the shared compound Poisson frailty model. See Wienke (2011) for more details.

4 Results

In this section we present and discuss empirical results obtained by fitting the models of Sect. 3 to the data described in Sect. 2. We began by estimating the Kaplan and Meier (1958) survival functions across the levels of the covariates for the entire country. The complement probabilities of these survival functions (that is, the probabilities of transition to parenthood) expressed in percentages are shown in Fig. 1 for region, residence, birth cohort, and education. We see that women from the two administrative cities (Addis Abeba and Dire Dawa) enter parenthood at much a slower rate than women in other regions. The same is true for women from urban areas in general and those with some education. For obvious reasons, the youngest cohorts have lower transition rates as they have not yet been exposed to the event long enough.

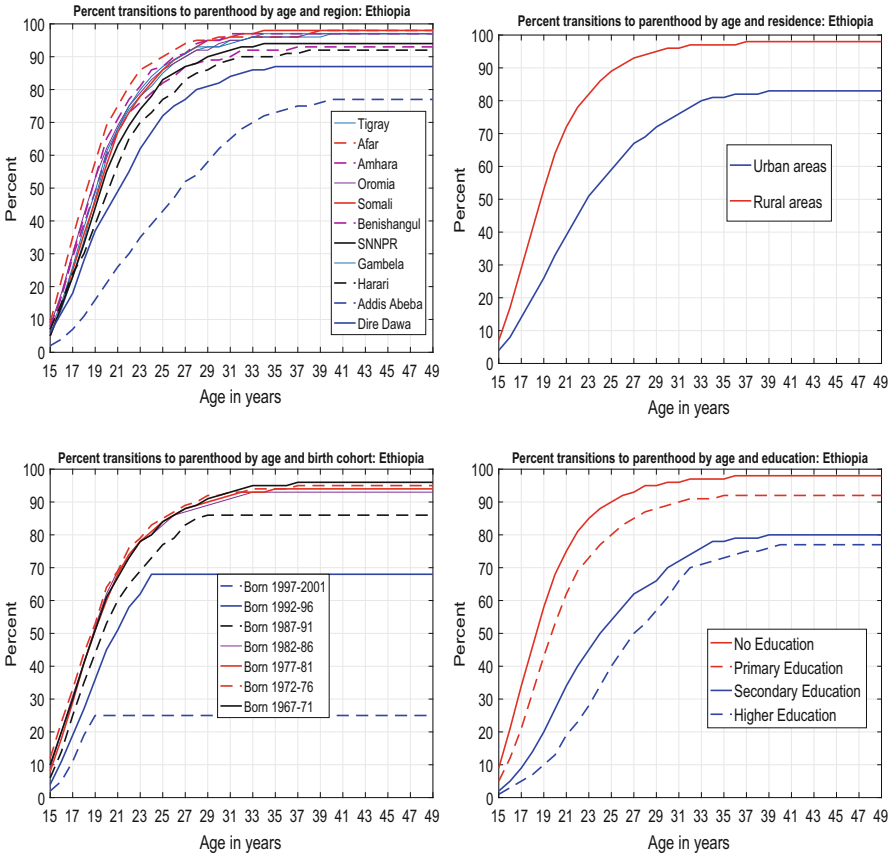


Fig. 1 Percentage transitions to parenthood by age and across regions, residence, birth cohort, and education

4.1 Results from Standard and Frailty Models for the Entire Country

The next step in our analysis was to fit the conventional Cox proportional hazards model (2), the Cox frailty model (5) with gamma distributed random components (7) and the Cox frailty model (5) with log-normal distributed random components (11) to data for the entire country. Region, residence, birth cohort, and education were included as covariates and the first level of each covariate was used as a baseline (reference) level.

Results from these three models are reported, in the form of relative intensities of transition to parenthood, in Table 3 together with frailty p-values, variances of the frailty terms and corresponding intra-household correlation (IHC) for the two models with frailties.

Table 3 Estimated relative intensities of transition to parenthood among Ethiopian women: Results from standard Cox PH model and Cox PH models with frailty

Covariate	Standard Cox PH	Cox PH with gamma-frailty	lognormal-frailty lognormal-frailty
Tigray (Ref.)	1	1	1
Afar	1.1189**	1.1194**	1.1244**
Amhara	0.8978**	0.8980**	0.9000**
Oromia	0.9710	0.9711	0.9725
Somali	0.8403***	0.8400***	0.8366***
Benishangul	1.0300	1.0304	1.0351
SNNPR	0.8122***	0.8119***	0.8084***
Gambela	1.1172**	1.1177**	1.1231**
Harari	0.9592	0.9593	0.9604
Addis Abeba	0.5725***	0.5721***	0.5667***
Dire Dawa	0.8608***	0.8608***	0.8614***
Urban (Ref.)	1	1	1
Rural	1.3654***	1.3654***	1.3666***
Born 1997–2001 (Ref.)	1	1	1
Born 1992–1996	1.8770***	1.8771***	1.8777***
Born 1987–1992	2.1702***	2.1707***	2.1775***
Born 1982–1986	2.4462***	2.4474***	2.4608***
Born 1977–1981	2.2497***	2.2503***	2.2585***
Born 1972–1976	2.3432***	2.3446***	2.3607***
Born 1967–1971	2.1269***	2.1278***	2.1380***
No Education (Ref.)	1	1	1
Primary	0.8493***	0.8490***	0.8449***
Secondary	0.4699***	0.4694***	0.4629***
Higher	0.3324***	0.3318***	0.3255***
Frailty p-value	–	0.39	0.13
Variance of random effects	–	0.00206	0.02840
IHC (Intra HH correlation)	–	0.001	0.014

* Estimate significant at 10% level of significance
** Estimate significant at 5% level of significance
*** Estimate significant at 1% level of significance

As would be expected, there are significant differentials in the intensity of transition to parenthood across regions, residence, birth cohort, and education. Further, the effects are in the expected direction with, for instance, a relative intensity of 0.57 for women from the capital (relative to Tigray region) and a relative intensity of 1.12 for women from the Afar and Gambela regions (12% higher intensity compared to women from the Tigray region). Rural women enter parenthood at a rate that is 37% higher than their urban counterparts while the educational gradient is negative where women with above secondary education having a relative intensity of transition to parenthood of 0.33 relative to those with

no education. These differentials across the levels of the four covariates did not change at all when we accounted for clustering of women into households. This is in accordance with the frailty estimates at the bottom of the table where the p-values are very high and the variances (and, hence, the corresponding intra-household correlations) very small.

It would, thus, be of interest to partition the data by region and fit the above three models to each of the nine regions and two city administrations.

4.2 Results from Stratified Cox Models with No Frailty

The final three steps in our analysis were to fit the conventional Cox proportional hazards models (2), the Cox frailty models (5) with gamma distributed random components (7), and the Cox frailty models (5) with log-normal distributed random components (11) separately to data for each of the nine regions and two city administrations.

Region is no longer a covariate in these stratified models and, hence, each model was fit with the three covariates: residence, birth cohort, and education. We have also fit the above three models for the entire country with these three covariates to examine if their effects change in models without region. Again, the first levels of each covariate were used as a baseline (reference) levels.

Results from conventional Cox proportional models (2) for each region are reported in the upper panel of Table 4. Estimates that are significant at 5 % level are shown with bold emphasis while those that are significant at 10 % are shown with italic emphasis. Entries across columns for a given row show differences in effects of a that row (covariate) across regions. Accordingly, we note that residential differences in transition to parenthood are significant for the entire country as well as in the Afar, Somali, Gambela, and Harari regions as well as in the Dire Dawa city administration. On the other hand, the residential difference was insignificant in Tigray, Amhara, Benishangul-Gumuz, and the SNNP regions and marginally significant in the Oromia region. Further, we find insignificant or weak effects of birth cohort in the Afar, Somali, and Harari regions. Finally, differentials across educational levels are significant in most regions but we find no difference between women with no education and those with primary education in Tigray, Afar, Somali, Benishangul-Gumuz, Gambela, Harari, Addis Abeba, and Dire Dawa. For Somali and Gambela regions there is no difference even between women with no education and those with secondary education.

4.3 Results from Stratified Cox Models with Gamma Frailty

Results from a Cox frailty model (5) with gamma distributed random components (7) for each region are reported in the middle panel of Table 4.

There is no appreciable change in any of the estimates or their significance compared to the corresponding values in the top panel. This is also reflected in the frailty-related estimates at the bottom of the middle panel. The estimates of the variances and, hence, the corresponding intra-household correlations (IHC) are very small. Further, the p-values associated with the household random effects are very large save those of the two city administrations, Addis Abeba and Dire Dawa where the p-values are 0.10 and 0.13, respectively.

4.4 Results from Stratified Cox Models with Log-Normal Frailty

Lastly, results from a Cox frailty model (5) with log-normal distributed random components (11) are shown at the bottom panel of Table 4.

Again, the estimates and their significance do not change to any appreciable extent from their corresponding values at the top panel of Table 4. But, we note some significant changes in the frailty-related estimates at the bottom of the table. Now, the household random effects are significant for the entire country and the two city administrations Addis Abeba and Dire Dawa with p-values of 0.04, 0.05, and 0.045, respectively.

5 Summary and Concluding Remarks

In this chapter we presented multilevel survival models that account for hierarchical structure of data. The specific application was regional differentials in the intensity of transition to parenthood among Ethiopian women. Since women were clustered within households we used the households as units of analysis and treated women in the same household as correlated cases (multi-levels) of the same observation. Such formulation enabled us to add household-specific unobserved heterogeneity (random effect) that affects the outcome at woman level.

We fitted the conventional Cox proportional intensities model as well as two of its shared frailty variants (with gamma distributed and log-normal distributed random effects, respectively) to data for the entire country and, separately, for each of its 11 administrative regions.

The results for this particular application showed that the introduction of household-level random effects made no appreciable difference in the estimation of observed covariate effects in most (nine) of the regions. On the other hand, we found significant household-level random effects for the entire country as well as in the two city administrations with inhabitants from diverse ethnic groups. More interestingly, these household-level random effects were sensitive to the choice of the distribution and were significant only when they were assumed to follow log-normal distribution.

Table 4 Estimated relative intensities of transition to parenthood among Ethiopian women: Results from standard Cox PH model and Cox PH models with frailty

Covariate	Ethiopia	Tigray	AfarR	Amhara	Oromia	Somali	Benish.	SNNPR	Gambela	Harari	AddisA	DireD
	Results from standard Cox proportional intensities models											
Rural	1.58	1.09	1.91	1.09	1.29	1.29	1.12	0.89	1.50	1.73	-	2.10
Born92-96	1.87	2.22	1.48	2.42	1.89	1.49	2.54	1.98	1.97	1.35	2.59	2.02
Born87-92	2.16	2.11	1.31	3.18	1.95	1.49	3.11	2.69	2.22	1.46	5.86	2.54
Born82-86	2.39	2.94	1.61	3.59	2.11	1.44	2.92	3.15	2.32	1.71	6.79	3.01
Born77-81	2.21	2.63	1.59	3.83	1.63	1.06	2.95	3.16	2.00	1.25	6.82	3.93
Born72-76	2.30	2.98	1.26	4.40	1.82	0.98	2.91	3.69	1.97	1.41	8.76	2.61
Born67-71	2.11	3.15	1.14	3.30	1.53	0.88	3.22	3.34	1.83	1.02	8.32	3.37
Primary	0.83	0.91	0.89	0.74	0.68	0.99	0.99	0.85	1.11	1.09	0.84	0.89
Second.	0.44	0.43	0.47	0.32	0.34	0.75	0.50	0.33	0.91	0.59	0.48	0.49
Higher	0.31	0.26	0.44	0.20	0.32	0.38	0.13	0.26	0.54	0.41	0.39	0.36
	Results from Cox proportional intensities models with gamma frailty											
Rural	1.58	1.09	2.00	1.09	1.29	1.29	1.12	0.89	1.50	1.74	-	2.18
Born92-96	1.87	2.23	1.51	2.42	1.89	1.49	2.55	1.98	1.97	1.35	2.62	2.00
Born87-92	2.16	2.12	1.32	3.19	1.95	1.49	3.13	2.69	2.22	1.46	6.04	2.58
Born82-86	2.39	2.95	1.65	3.61	2.12	1.44	2.95	3.16	2.33	1.72	6.97	3.05
Born77-81	2.21	2.63	1.64	3.84	1.63	1.06	2.97	3.17	2.00	1.25	7.17	4.08
Born72-76	2.30	2.99	1.24	4.42	1.82	0.98	2.93	3.71	1.98	1.41	9.24	2.68
Born67-71	2.11	3.17	1.15	3.30	1.52	0.87	3.25	3.35	1.83	1.02	8.90	3.54
Primary	0.83	0.91	0.89	0.74	0.68	0.99	0.99	0.85	1.12	1.09	0.83	0.89
Second.	0.44	0.43	0.45	0.32	0.34	0.75	0.50	0.33	0.91	0.59	0.45	0.46
Higher	0.31	0.26	0.41	0.20	0.32	0.38	0.13	0.26	0.54	0.41	0.37	0.34
Frailty p-value	0.35	0.34	0.18	0.37	0.54	0.32	0.33	0.86	0.60	0.31	0.10	0.13
Var(randomeff.)	0.003	0.002	0.082	0.013	0.000	0.001	0.005	0.000	0.000	0.002	0.097	0.085
IHC	0.001	0.001	0.039	0.007	0	0.001	0.003	0	0	0.001	0.046	0.041
	Results from Cox proportional intensities models with log-normal frailty											
Rural	1.58	1.10	1.98	1.09	1.29	1.30	1.13	0.89	1.50	1.78	-	2.20
Born92-96	1.88	2.23	1.50	2.41	1.89	1.49	2.58	1.98	1.97	1.37	2.63	1.99
Born87-92	2.17	2.12	1.32	3.19	1.95	1.50	3.24	2.69	2.22	1.50	6.07	2.59
Born82-86	2.41	2.95	1.64	3.61	2.12	1.44	3.10	3.16	2.34	1.78	7.00	3.06
Born77-81	2.22	2.68	1.63	3.83	1.63	1.06	3.03	3.17	2.00	1.26	7.21	4.14
Born72-76	2.32	3.03	1.24	4.43	1.82	0.98	3.06	3.71	1.98	1.46	9.27	2.71
Born67-71	2.12	3.21	1.15	3.29	1.52	0.87	3.40	3.35	1.84	1.02	9.00	3.61
Primary	0.82	0.90	0.89	0.73	0.68	0.99	1.02	0.85	1.12	1.10	0.82	0.89
Second.	0.44	0.42	0.46	0.31	0.34	0.75	0.50	0.33	0.91	0.58	0.45	0.46
Higher	0.31	0.25	0.42	0.19	0.31	0.38	0.12	0.26	0.54	0.40	0.37	0.34
Frailty p-value	0.04	0.27	0.22	0.32	0.39	0.40	0.097	0.35	0.39	0.27	0.051	0.045
Var(randomeff.)	0.040	0.038	0.054	0.026	0.003	0.012	0.087	0.001	0.005	0.059	0.126	0.133
IHC	0.019	0.019	0.026	0.013	0.002	0.006	0.042	0.001	0.003	0.029	0.059	0.062

Another interesting result when analyzing the data for the entire country relates to significance of the household-level random effects. These were insignificant (for both gamma and log-normal distributions) when region was included as a covariate in the models. But, they turned out to be significant under the log-normal distribution when region was excluded from the model. This should not be surprising because, as stated before, the effect of the household random effects was significant in two of the regions. This effect is expected to disappear if region is included as a covariate in the model because it will be absorbed in the effects of region.

The absence of household random effects in nine of the regions should not be taken as a green light to ignore clustering of women into households. In the multilevel (frailty) models we used in this chapter, the frailty term is a measure of the relative intensity that women within a household share. Conditional on the frailty term the survival times of women in that household were assumed to be independent of each other.

On the other hand, the definition of household is not clear in Demographic and Health Surveys. As we saw in Table 2, the number of women within a household varied between 1 and 9 in the Afar region and between 1 and 15 in the capital, Addis Abeba. Whether household means women living under the same roof or an extended family or clan it is most likely that survival times of women in a household are correlated but not necessarily shared. In such cases a correlated frailty model which is a natural extension of the shared frailty model may be more appropriate because it allows inclusion of additional correlation parameters to address associations between event times of women within the same household. If these correlations equal to one, the shared frailty model is obtained as a special case of the correlated frailty model.

Future works in the area, which we intend to pursue, may therefore explore if other frailty models capture effects of household random effects more accurately. These include the nested, joint, and additive frailty models in Rondeau et al. (2012, 2021) and the four variants of gamma frailty models in Martins et al. (2019).

Acknowledgments The work on this chapter was initiated as a project for a senior undergraduate essay in Statistics completed at the Department of Statistics, Stockholm University, Sweden—see Akinyi Lagehäll and Yemane (2021). We would like to thank Carren Melinder for comments on draft versions of that essay. The data analyzed in this chapter was generously provided by the DHS program (https://www.dhsprogram.com).

Appendix: R-Codes Used for Computing the Results Reported in the Tables of This Chapter

```
#####################################################################
# Importing the data set into R and naming it ethiopia16
library(readxl)
ethiopia16 <- read_excel("E:/GG-New-Data-with-Cum-HH.xlsx")
View(ethiopia16)
```

```
#####################################################################
# Creating subsets for each region
Tigray <- subset(ethiopia16, Region == 1)
Afar <- subset(ethiopia16, Region == 2)
Amhara <- subset(ethiopia16, Region == 3)
Oromia <- subset(ethiopia16, Region == 4)
Somali <- subset(ethiopia16, Region == 5)
Benishangul <- subset(ethiopia16, Region == 6)
SNNPR <- subset(ethiopia16, Region == 7)
Gambela <- subset(ethiopia16, Region == 8)
Harari <- subset(ethiopia16, Region == 9)
Addis_Ababa <- subset(ethiopia16, Region == 10)
Dire_Dawa <- subset(ethiopia16, Region == 11)
#####################################################################
## 1. Kaplan-Meier estimated survival curves by region ##
#####################################################################
library(survival)
RegionTigray <- which(ethiopia16$Region==1)
regionT <- ethiopia16[RegionTigray,]
km.regionT <-survfit(Surv(Exposure, Status)~1, data=regionT)
RegionAfar <- which(ethiopia16$Region==2)
regionAf <- ethiopia16[RegionAfar,]
km.regionAf <-survfit(Surv(Exposure, Status)~1, data=regionAf)
RegionAmhara <- which(ethiopia16$Region==3)
regionAm <- ethiopia16[RegionAmhara,]
km.regionAm <-survfit(Surv(Exposure, Status)~1, data=regionAm)
RegionOromia <- which(ethiopia16$Region==4)
regionO <- ethiopia16[RegionOromia,]
km.regionO <-survfit(Surv(Exposure, Status)~1, data=regionO)
RegionSomali <- which(ethiopia16$Region==5)
regionS <- ethiopia16[RegionSomali,]
km.regionS <-survfit(Surv(Exposure, Status)~1, data=regionS)
RegionBenishangul <- which(ethiopia16$Region==6)
regionB <- ethiopia16[RegionBenishangul,]
km.regionB <-survfit(Surv(Exposure, Status)~1, data=regionB)
RegionSNNPR <- which(ethiopia16$Region==7)
regionSN <- ethiopia16[RegionSNNPR,]
km.regionSN <-survfit(Surv(Exposure, Status)~1, data=regionSN)
RegionGambela <- which(ethiopia16$Region==8)
regionG <- ethiopia16[RegionGambela,]
km.regionG <-survfit(Surv(Exposure, Status)~1, data=regionG)
RegionHarari <- which(ethiopia16$Region==9)
regionH <- ethiopia16[RegionHarari,]
km.regionH <-survfit(Surv(Exposure, Status)~1, data=regionH)
RegionAddis <- which(ethiopia16$Region==10)
regionAA <- ethiopia16[RegionAddis,]
km.regionAA <-survfit(Surv(Exposure, Status)~1, data=regionAA)
RegionDire <- which(ethiopia16$Region==11)
regionDD <- ethiopia16[RegionDire,]
km.regionDD <-survfit(Surv(Exposure, Status)~1, data=regionDD)
plot(km.regionDD, main="Figure␣1:␣Survival␣functions␣by␣region:
KM␣estimation", col = "green",
cex.axis=0.7, xlab = "Exposure␣in␣months", ylab = "Survival
functions", conf.int = "none")
```

```
lines(km.regionAA, col = "orange", conf.int = "none")
lines(km.regionH, col = "darkblue", conf.int = "none")
lines(km.regionG, col = "yellow", conf.int = "none")
lines(km.regionSN, col = "red", conf.int = "none")
lines(km.regionB, col = "blue", conf.int = "none")
lines(km.regionS, col = "black", conf.int = "none")
lines(km.regionO, col = "purple", conf.int = "none")
lines(km.regionAm, col= "pink", conf.int = "none")
lines(km.regionAf, col="violet", conf.int = "none")
lines(km.regionT, col = "brown", conf.int = "none")
legend("topright", c("Dire␣Dawa", "Addis␣Ababa", "Harari",
"Gambela", "SNNPR", "Benishangul", "Somali", "Oromia", "Amhara",
"Afar", "Tigray"), lty = c(1:1), cex=0.7, col = c("green",
"orange", "darkblue", "yellow", "red", "blue", "black", "purple",
"pink", "violet", "brown"))
#####################################################################
## 2. COX proportional models for the whole country and for each
region ##
#####################################################################
#
coxall <- coxph(Surv(Exposure, Status) ~ as.factor(Region)
+ as.factor(Residence) + as.factor(Cohort) + as.factor(Education),
method="breslow", data = ethiopia16)
summary(coxall)
#
coxTigray <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Tigray)
summary(coxTigray)
coxAfar <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Afar)
summary(coxAfar)
coxAmhara <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Amhara)
summary(coxAmhara)
coxOromia <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Oromia)
summary(coxOromia)
coxSomali <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Somali)
summary(coxSomali)
coxBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor
 (Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Benishangul)
summary(coxBenishangul)
coxSNNPR <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) , method="breslow",
data = SNNPR)
summary(coxSNNPR)
```

```
coxGambela <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) , method="breslow",
data = Gambela)
summary(coxGambela)
coxHarari <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Harari)
summary(coxHarari)
coxAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor(Cohort)
+ as.factor(Education), method="breslow", data = Addis_Ababa)
summary(coxAddis_Ababa)
coxDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = Dire_Dawa)
summary(coxDire_Dawa)
#####################################################################
## 3. Testing the proportional hazards assumptions ##
#####################################################################
cox.zph(coxall)
cox.zph(coxTigray)
cox.zph(coxAfar)
cox.zph(coxAmhara)
cox.zph(coxOromia)
cox.zph(coxSomali)
cox.zph(coxBenishangul)
cox.zph(coxSNNPR)
cox.zph(coxGambela)
cox.zph(coxHarari)
cox.zph(coxAddis_Ababa)
cox.zph(coxDire_Dawa)
#####################################################################
## 4. Frailty models for the whole country and for each region ##
#####################################################################
#Gamma- and log-normal frailty distributed models
#Ethiopia
#
GammaAll <- coxph(Surv(Exposure, Status) ~ as.factor(Region)
+ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), method="breslow", data = ethiopia16)
summary(GammaAll)
#
# The command below was updated by GG (with cumulative HH ranging
b/n 1 and 4722)
#
lognormalAll <- coxph(Surv(Exposure, Status) ~ as.factor(Region)
+ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), method="breslow", data = ethiopia16)
summary(lognormalAll)
#
#Tigray
GammaTigray <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
```

```
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Tigray)
summary(GammaTigray)
lognormTigray <- coxph(Surv(Exposure, Status) ~ as.factor
(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Tigray)
summary(lognormTigray)
#Afar
GammaAfar <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Afar)
summary(GammaAfar)
lognormAfar <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Afar)
summary(lognormAfar)
#Amhara
GammaAmhara <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Amhara)
summary(GammaAmhara)
lognormAmhara <- coxph(Surv(Exposure, Status) ~ as.factor\
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Amhara)
summary(lognormAmhara)
#Oromia
GammaOromia <-coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Oromia)
summary(GammaOromia)
lognormOromia <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Oromia)
summary(lognormOromia)
#Somali
GammaSomali <-coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Somali)
summary(GammaSomali)
lognormSomali <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Somali)
summary(lognormSomali)
#Benishangul
GammaBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort)
+ as.factor(Education) + frailty(HouseHold,distribution="gamma"),
```

```
data=Benishangul)
summary(GammaBenishangul)
lognormBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort)
+ as.factor(Education) + frailty(HouseHold,distribution=
  "gaussian"),
data=Benishangul)
summary(lognormBenishangul)
#SNNPR
GammaSNNPR <-coxph(Surv(Exposure, Status) ~ as.factor
(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=SNNPR)
summary(GammaSNNPR)
lognormSNNPR <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=SNNPR)
summary(lognormSNNPR)
#Gambela
GammaGambela <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Gambela)
summary(GammaGambela)
lognormGambela <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Gambela)
summary(lognormGambela)
#Harari
GammaHarari <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Harari)
summary(GammaHarari)
lognormHarari <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Harari)
summary(lognormHarari)
#Addis Ababa
GammaAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor
  (Cohort)
+ as.factor(Education) + frailty(HouseHold,distribution="gamma"),
data=Addis_Ababa)
summary(GammaAddis_Ababa)
lognormAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor
  (Cohort)
+ as.factor(Education) + frailty(HouseHold,distribution="gaussian"),
data=Addis_Ababa)
summary(lognormAddis_Ababa)
#Dire Dawa
GammaDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor
```

```
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), data=Dire_Dawa)
summary(GammaDire_Dawa)
lognormDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor
  (Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gaussian"), data=Dire_Dawa)
summary(lognormDire_Dawa)
#################################################################
## 5. The different frequencies ##
#################################################################
# Ethiopia
table(ethiopia16$Residence)
table(ethiopia16$Cohort)
table(ethiopia16$Education)
table(ethiopia16$Residence,ethiopia16$Status)
table(ethiopia16$Cohort,ethiopia16$Status)
table(ethiopia16$Education,ethiopia16$Status)
#Tigray
table(Tigray$Residence)
table(Tigray$Cohort)
table(Tigray$Education)
table(Tigray$Residence,Tigray$Status)
table(Tigray$Cohort,Tigray$Status)
table(Tigray$Education,Tigray$Status)
#Afar
table(Afar$Residence)
table(Afar$Cohort)
table(Afar$Education)
table(Afar$Residence,Afar$Status)
table(Afar$Cohort,Afar$Status)
table(Afar$Education,Afar$Status)
#Amhara
table(Amhara$Residence)
table(Amhara$Cohort)
table(Amhara$Education)
table(Amhara$Residence,Amhara$Status)
table(Amhara$Cohort,Amhara$Status)
table(Amhara$Education,Amhara$Status)
#Oromia
table(Oromia$Residence)
table(Oromia$Cohort)
table(Oromia$Education)
table(Oromia$Residence,Oromia$Status)
table(Oromia$Cohort,Oromia$Status)
table(Oromia$Education,Oromia$Status)
#Somali
table(Somali$Residence)
table(Somali$Cohort)
table(Somali$Education)
table(Somali$Residence,Somali$Status)
table(Somali$Cohort,Somali$Status)
table(Somali$Education,Somali$Status)
```

```
#Benishangul
table(Benishangul$Residence)
table(Benishangul$Cohort)
table(Benishangul$Education)
table(Benishangul$Residence,Benishangul$Status)
table(Benishangul$Cohort,Benishangul$Status)
table(Benishangul$Education,Benishangul$Status)
#SNNPR
table(SNNPR$Residence)
table(SNNPR$Cohort)
table(SNNPR$Education)
table(SNNPR$Residence,SNNPR$Status)
table(SNNPR$Cohort,SNNPR$Status)
table(SNNPR$Education,SNNPR$Status)
#Gambela
table(Gambela$Residence)
table(Gambela$Cohort)
table(Gambela$Education)
table(Gambela$Residence,Gambela$Status)
table(Gambela$Cohort,Gambela$Status)
table(Gambela$Education,Gambela$Status)
#Harari
table(Harari$Residence)
table(Harari$Cohort)
table(Harari$Education)
table(Harari$Residence,Harari$Status)
table(Harari$Cohort,Harari$Status)
table(Harari$Education,Harari$Status)
#Addis Ababa
table(Addis_Ababa$Residence)
table(Addis_Ababa$Cohort)
table(Addis_Ababa$Education)
table(Addis_Ababa$Residence,Addis_Ababa$Status)
table(Addis_Ababa$Cohort,Addis_Ababa$Status)
table(Addis_Ababa$Education,Addis_Ababa$Status)
#Dire Dawa
table(Dire_Dawa$Residence)
table(Dire_Dawa$Cohort)
table(Dire_Dawa$Education)
table(Dire_Dawa$Residence,Dire_Dawa$Status)
table(Dire_Dawa$Cohort,Dire_Dawa$Status)
table(Dire_Dawa$Education,Dire_Dawa$Status)
length(unique(ethiopia16$HouseHold)) #487
length(unique(Tigray$HouseHold)) #452
length(unique(Afar$HouseHold)) #410
length(unique(Amhara$HouseHold)) #452
length(unique(Oromia$HouseHold)) #466
length(unique(Somali$HouseHold)) #451
length(unique(Benishangul$HouseHold)) #415
length(unique(SNNPR$HouseHold)) #459
length(unique(Gambela$HouseHold)) #390
length(unique(Harari$HouseHold)) #383
length(unique(Addis_Ababa$HouseHold)) #442
length(unique(Dire_Dawa$HouseHold)) #402
```

```
#
############################################################
6. Additional by GG (three models for entire country without
region)
############################################################
#
GG_Coxall <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education), method="breslow",
data = ethiopia16)
summary(GG_Coxall)
#
GG_GammaAll <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)
+ as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,
distribution="gamma"), method="breslow", data = ethiopia16)
summary(GG_GammaAll)
#
GG_lognormalAll <- coxph(Surv(Exposure, Status)
~ as.factor(Residence) + as.factor(Cohort) + as.factor(Education)
+ frailty(HouseHold, distribution="gaussian"), method="breslow",
data = ethiopia16)
summary(GG_lognormalAll)
#
################################################################
```

References

Akinyi Lagehäll, A. & Yemane, E. (2021). *Multilevel Cox Regression of Transition to Parenthood among Ethiopian Women*. Bachelor Thesis 2021:2, Department of Statistics, Stockholm University.

Austin, P. C. (2017). A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review, 85*(2), 185–203.

Bolstad, W. M., & Manda, S. O. (2001). Investigating child mortality in Malawi using family and community random effects: a Bayesian analysis. *Journal of the American Statistical Association, 96*(453), 12–19.

Central-Statistical-Agency-Ethiopia, C. S. A. and ICF, U.S.A. (2016). *Demographic and health survey 2016*. Ethiopia: Central Statistical Agency (CSA).

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika, 65*(1), 141–151.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–202.

Cox, D. R. (1975). Partial likelihood. *Biometrika, 62*(2), 269–276.

Guo, G., & Rodriguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. *Journal of the American Statistical Association, 87*(420), 969–976.

Heckman, J. J., & Singer, B. (1982). Population heterogeneity in demographic models. In *Multidimensional mathematical demography* (pp. 567–599). Elsevier.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society,* 271–320.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481.

Kendall, M. G. (1962). *Rank correlation methods, 3rd ed.*. Griffin.

Lancaster, T. (1990). *The econometric analysis of transition data*, vol. 17. Cambridge University Press.

Martins, A., Aerts, M., Hens, N., Wienke, A., & Abrams, S. (2019). Correlated gamma frailty models for bivariate survival time data. *Statistical Methods in Medical Research, 28*(10–11), 3437–3450.

McGilchrist, C. A. (1993). Reml estimation for survival models with frailty. *Biometrics, 49*(1), 221–225.

McGilchrist, C., & Aisbett, C. (1991). Regression with frailty in survival analysis. *Biometrics* **47**(2), 461–466.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B (Methodological), 44*(3), 414–422.

Rondeau, V., Mazroui, Y., & Gonzalez, J. R. (2012). frailtypack: an r package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software, 47*(4), 1–28.

Rondeau, V., Gonzalez, J. R., Yassin Mazroui, A., Mauguen, A. D., Laurent, A., Lopez, M., Krol, A., Sofeu, C. L., Dumerc, J., Rustand, D. et al. (2021). Package 'frailtypack'.

Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association, 92*(438), 426–435.

Therneau, T. M., & Lumley, T. (2021). Package 'survival' version 3.2-11. *Survival Analysis Published on CRAN, 3.2*(11), 1–185.

Therneau, T. M., Grambsch, P. M., & Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics, 12*(1), 156–175.

Trussell, J., & Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. *Sociological Methodology, 15*, 242–276.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*(3), 439–454.

Wienke, A. (2011). *Frailty models in survival analysis*. CRC Press.

Application of Multiple Imputation, Inverse Probability Weighting, and Double Robustness in Determining Blood Donor Deferral Characteristics in Malawi

Evaristar N. Kudowa and Mavuto F. Mukaka

Abstract Missing data occur in most epidemiological studies and may reduce internal validity of study findings. Biased and inefficient estimates may result if missing values are not properly handled during analysis. Application of Complete case (CC) analysis to handle missing data may produce biased estimates if data are not missing completely at random. We, therefore, need principled methods to address missing data. We addressed missing data for blood donor retrospective cohort to estimate predictors of donor deferral using Logistic regression model. Multiple Imputation (MI), Inverse Probability Weighting (IPW), and Double Robustness (DR-IPW) were applied to correct for missingness. CC estimates had wider confidence intervals, consistent with IPW estimates. MI and DR produced narrow confidence intervals relative to CC and IPW methods. MI indicated higher odds of deferral among syphilis-infected donors OR: 1.24 (95% CI: 1.05, 1.48) and lower odds of deferral among donors with higher number of blood donations OR: 0.90 (95% CI: 0.82, 0.98). Estimates for MI and DR were precise compared to CC and DR methods. DR guards against model misspecification making it the preferred method for more accurate analysis but is limited by its unavailability in most existing software. Instead, MI which is widely available can be used.

Keywords Multiple Imputation · Inverse Probability · Double Robustness · Complete case · Blood donation · Missing data · Missing values · Ad-hoc methods · Univariate missingness · Intermittent missingness · Monotonic missingness · Missingness mechanism · Random missingness · Non-random

E. N. Kudowa (✉)
University of North Carolina Project, Lilongwe, Malawi

University of Malawi, Chancellor College, Zomba, Malawi

M. F. Mukaka
Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2_18

missingness · Non-ignorable · Sensitivity analysis · Unbiased estimates · Biased estimates · Logistic regression · Semiparametric · Fully observed · Consistent estimates · Efficient estimates · Inefficient estimates · Incomplete values · Randomness · Pooled estimates · Rubin rule · Imputations · Observed values · MICE algorithm

1 Introduction

Due to high clinical demand of blood supply, the World Health Organization recommends blood donation systems to collect blood from voluntary donors. Blood donors are screened before donation and ineligible donors are deferred for donation. In trying to determine the characteristics of deferred donors, we observed a high proportion of missing values (96%) in both the predictor variable (number of donations) and the outcome variable (deferral status). These missing data could potentially lead to biased estimates and affect interpretation of study findings if not properly corrected during analysis. To address analysis challenges introduced by missing data, many researchers apply complete case or list wise deletion methods, which excludes observations with incomplete information from the analysis. Ad-hoc methods such as complete case and available case analyses are the choice of methods for addressing missing data only because of their programming convenience (Carpenter et al., 2006). Despite the ease of these methods when applied to incomplete data analysis, they only give valid estimates when data are missing completely at random and not necessarily when they are missing at random. However, their estimates may be less efficient due to the reduction in sample size (Eekhout et al., 2012). Additionally, ad-hoc methods do not utilize statistical principles for handling missing values and, therefore, do not adequately address issues raised by missing data (Carpenter et al., 2006).

Many scientific studies that are prone to missing data do not state the methods used to handle missing values in their analysis. A review of studies that had missing data showed that in 262 of the studies that had missing data, 46% of the studies could not clearly define the type of missing data. There were 81% studies that performed a complete case analysis and 14% used a single imputation technique, such as mean imputation, single regression, or last observation carried forward. More advanced methods, such as multiple imputation (MI), maximum likelihood estimation, and inverse probability weighting, were reported in 8%, 2%, and 3% of the studies, respectively (Eekhout et al., 2012).

However, the impact of missing data on study results depends on the degree or percentage of missing data. Statistical results obtained when missing data are less than 5% are typically insensitive. For missing data that are between 5 and 15% the impact depends on the context or study type, whereas results obtained for missing data which is more than 15% are often sensitive and appropriate correction methods for missing data need to be applied to such data in order to obtain unbiased estimates (Harrell, 2001).

In this analysis we explore the application of multiple imputation, inverse probability of censoring weights and double robustness in correcting for missing data and compare their estimates to the estimates obtained from complete case analysis.

2 Missing Data Patterns

This refers to the order of missing values in the dataset, which can be univariate, intermittent, monotonic, or arbitrary. With univariate missing data pattern, missing data values are only on one variable in the dataset and the rest are fully observed. Univariate missing data usually happens due to the study design. Monotonic (terminal) pattern occurs when the data values are observed from baseline until drop-out and no subsequent information is collected. Little and Rubin (2020) proposed that if we let $Y = [y_{ij}]$, where $i = 1, \ldots, n$ and $j = 1, \ldots, k$ denote an $n \times k$ completely observed dataset, where the value y_{ij} represents the value Y_j for subject i. For the missingness to be monotonic , whenever y_{ij} is missing, y_{ik} becomes missing $\forall k > j$.

Monotonic missing data pattern is mostly encountered in studies of a population with high mortality and morbidity rate and longitudinal studies which are more prone to drop-outs or attrition (Little & Rubin, 2020). With monotonic pattern ordering of data units during analysis is necessary for observing the pattern in the data (Gordon, 2010). In intermittent missing data pattern, the missingness occurs between successful or observed assessments. Under this pattern, subjects that miss an observation at some point return to the study afterwards. This is mostly encountered in studies of individuals with chronic conditions such as cancer, asthma, or diabetes. Mixed (Arbitrary) pattern involves both monotonic and intermittent missing patterns. This is when a period of intermittent missingness is followed by monotone missingness. Unlike the other patterns, an arbitrary pattern is computationally difficult to handle in practice (Dong and Peng, 2013).

3 Missing Data Nomenclature

Missing data nomenclature are mechanisms that characterize the reasons for the missing data. These missing data mechanisms/response mechanisms were first described by Donald Rubin in 1976 and were classified into three categories. Rubin (1976) indicated the likelihood for every data point to have missing value and these likelihood can be managed through the missing data mechanisms. Understanding these mechanisms is crucial in selecting an appropriate analysis method, because the choice of missing data handling technique is dependent on the assumptions made (Baraldi & Enders, 2010).

Watanabe and Yamaguchi (2004) indicated the need to first understand reasons behind the missingness when interpreting results that have missing values. When this information is unknown, methods suitable for analyzing such cases should be used. If there are known reasons behind the missingness, then it is appropriate to firstly decide whether to ignore or take the mechanism into account. Ignoring the missingness mechanism is appropriate if the missing values are independent of the observed values, i.e., missingness due to sampling mechanism.

To define these missing data mechanisms we assume we have a random variable X having the values x^O if observed or x^m if missing and we let $R = 1$ if X is missing (x^m) with $P(R = 1)$ being the probability of missing X, $R = 0$ if X is observed (x^O) and $P(R = 0)$ is the probability of observing X. We also have a vector Z of all measured variables including X, which may also have missing data and another vector U for all unmeasured variables. Missing at random (MAR) occurs if after conditioning for unobserved data, the missingness depends on the observed data, i.e.,

$$P(R = 1|Z^O, Z^m, u) = P(R = 1|Z^O) \tag{1}$$

It is, however, difficult to verify MAR assumption using observed data (Yuan, 2014); therefore, sensitivity analysis ought to be made to verify MAR inferences (National Research Council, 2010). Missing not at random (MNAR) occurs when the missingness relates to the unobserved data after taking the observed values into account, i.e.,

$$P(R = 1|Z^O, Z^m, u) = P(R = 1|Z^O, Z^m, u) \tag{2}$$

This implies that the chance of seeing Z depends on Z^m, even after conditioning on Z^O. Thus, there is a relationship between what would have been observed and the missingness. MNAR can occur if subjects refuse to reveal something very personal about themselves, i.e., higher income individuals being less likely to reveal their income in a survey than individuals with low income. This mechanism is non-ignorable and the researcher is required to specify the missingness mechanism in the analysis model to achieve unbiased estimates. Missing completely at random (MCAR) is rare assumptions in real life and the missingness occurs by chance only. With MCAR there is equal probability of being missing for all cases (van Buuren, 2012). This occurs when the probability that the data are missing is not related to either the specific value that is supposed to be obtained or the set of observed values, i.e.,

$$P(R = 1|Z^O, Z^m, u) = P(R = 1) \tag{3}$$

The statistical advantage of data that are MCAR is that the estimated parameters remain unbiased despite power being lost in the design (Kang, 2013). This is because the observed values are actually a random subsample of the full dataset.

In addressing missing data problem, the study aimed at applying MI, IPW, and DR methods. Estimates from DR method were more precise compared to complete case and IPW methods but were similar to MI. DR and MI managed to address the missing data problem in our analysis.

4 Methods

4.1 Study Design and Population

This was a retrospective cohort study which reviewed regular blood donor records whose last donation date was between January 2014 and August 2015. We defined a regular donor as any person with at least two successful donations before the last donation date of the study period.

4.2 Data Sources, Description, and Management

We used secondary data from the Malawi Blood Transfusion Service (MBTS). Established in 2003, MBTS provides two-thirds of all blood used in hospitals with the other third being covered by hospital-based system. MBTS is a government (public) entity that ensures adequate supply of safe blood products to all patients in need of blood supply. In 2000, the World Health Organization launched a blood safety initiative of ensuring blood collection from non-remunerated (volunteered) donors who were identified as low risk donors compared to the then paid donors. Based on this initiative MBTS uses mobile services to routinely collect blood from various institutions from non-remunerated donors. Blood donor details are routinely collected and recorded in blood bank access database during blood donation with each donor uniquely identified. The collected details include social and demographic characteristics for the blood donors as well as their medical history.

4.3 Study Outcomes

The outcome variable was deferral status, a binary variable categorized into active and deferred donor, where deferred donor includes both permanent and temporary deferrals. The predictors for donor status include age, sex, number of donations, marital status, duration between successive donations, and hemoglobin (HB) levels. The other predictors include medical results for condition such as syphilis, HIV, HB levels, anemia, HBV, skin rashes, and Rhesus factor. But for this study sex,

number of donations, marital status, HIV status, and Syphilis status were identified as potential factors for determining deferral status. This selection was based on literature and availability of data. The dataset was not rich with information for the other known predictors for donor deferral.

4.4 Statistical Software and Methods

The statistical analyses were performed using STATA 14.1. We looked at distribution of blood donor demographic characteristics and deferral status using chi-square test. MI, IPW, and DR were applied to the data in order to achieve complete dataset or to correct for the missing values. Logistic regression analysis was used to model deferral characteristics among blood donors, firstly using complete case analysis and then using the imputed and weighted analysis methods. The fitted multivariable logistic regression model was given as:

$$Ln\left(\frac{\widehat{p}}{1-\widehat{p}}\right) = \beta_0 + \beta_1 \cdot \mathrm{m} + \beta_2 \cdot \mathrm{sex} + \beta_3 \cdot \mathrm{bd} + \beta_4 \cdot \mathrm{hiv} + \beta 5 \cdot \mathrm{sy} \qquad (4)$$

where m =married, bd =number of blood donations, sy =syphilis, and $\widehat{p}$ is the expected probability of being deferred for donation.

4.5 Assumption for Missingness Mechanism

MAR is the most plausible missing data assumption in real life because with this assumption the missingness depends on the observed data (Sterne et al., 2009). The study made an assumption of the missingness to be MAR and applied method for addressing missing data based on this assumption. T-test was performed between the indictor variable for missing deferral status and number of donations with sex variable. There was significant deference between those with observed deferral status and those with missing deferral status, thus the mean sex score was found to be significantly higher among those with missing deferral status indicating sex as a potential correlate of missingness. Variable sex was included in this model to satisfy MAR assumption. We discuss below the details of the methods of handling missing data methods that we applied, namely: multiple imputation, inverse probability weighting, and double robustness.

4.6 Multiple Imputation

This study used Multivariate Imputation by Chained Equations (MICE) to impute the missing values with more than 5% of missingness (Table 1). Ten imputations were made resulting in ten complete datasets and Rubin rule was then applied to the ten complete datasets to obtain pooled estimates.

MICE also known as Fully Conditional Specification (FCS) operates under MAR assumption and can handle variables of various type. It is semi-parametric and acts as an alternative to Joint Modelling (JM) when no suitable multivariate distribution can be found. FCS is flexible and easy to apply compared to JM and it accommodates an arbitrary missing data pattern. According to the expression by van Buuren and Groothuis-Oudshoorn (2011), we let Y the hypothetically complete data be partially observed random sample from a p-variate- multivariate distribution $P(Y|\theta)$. Here we assume that the multivariate distribution of Y is completely specified by a vector of unknown parameters θ. It is expected that the multivariate distribution of θ is explicitly known, even though knowing this distribution is quite challenging. van Buuren and Groothuis-Oudshoorn (2011), however, demonstrated that the MICE algorithm obtains the posterior distribution of θ through iterative sampling from condition distributions, i.e.,

$$\begin{aligned} &P(Y_1|Y_{-1}\theta_1) \\ &\vdots \\ &P(Y_p|Y_{-p}\theta_p) \end{aligned} \tag{5}$$

The parameters θ_1, θ_p obtained are specific to the respective conditional densities and are not necessarily the product of a factorization of the true joint distribution $P(Y|\theta)$. Due to the difficulty with direct sampling, the method uses Gibbs sampling which is a form of Markov Chain Monte Carlo algorithm. van Buuren and Groothuis-Oudshoorn (2011) presented the tth iteration of chained equations as a Gibbs sampler that successively draws:

$$\begin{aligned} \theta_1^{*(t)} &\sim P\left(\theta_1|Y_1^{obs}, Y_2^{t-1}, \ldots Y_p^{t-1}\right) \\ Y_1^{*(t)} &\sim P\left(Y_1|Y_1^{obs}, Y_2^{t-1}, \ldots Y_p^{t-1}, \theta_1^{*(t)}\right) \\ &\vdots \\ \theta_p^{*(t)} &\sim P\left(\theta_p|Y_p^{obs}, Y_1^{t-1}, \ldots Y_{p-1}^{t}\right) \\ Yp^{*(t)} &\sim P\left(Y_p|Y_1^{obs}, Y_1^{t}, \ldots Y_p\theta_1^{*(t)}\right) \end{aligned} \tag{6}$$

Each successive iteration results in generating an imputed variable where, $Y_j^{(t)} = \left(Y_j^{obs}, Y_j^{*(t)}\right)$ is the j^{th} imputed variable at the t^{th} iteration. This process then results in executing m parallel streams which generates an imputed dataset.

4.7 Inverse Probability Weighting

An indicator variable for missingness was generated based on whether the number of donations or deferral status was missing or not. Logistic regression was fitted on this indicator variable and other non-missing variables in the model of interest. This model indicates how the probability of being a complete record depends on the fully observed variables. Fitted probabilities of being a complete record were then obtained based on this model. To obtain weights for each observation an inverse of the fitted probabilities was calculated. Logistic regression of deferral status on the predictors was fitted and the estimated weights were passed to this model as an inverse of the fitted probabilities. In this case, we created a pseudo-population that would be observed had there been no missing values in our data.

IPW assumes consistent parameters if the probability of response π_i is known, regardless of the missingness mechanism. This still poses a challenge as π_i is usually not known. IPW involves specifying our model of interest usually a regression model relating the outcome and covariates. We model π_i (as a function of fully observed variables O_j), i.e., the model (logistic regression) for the probability of X_i being missing. The fitted model yields the predicted probabilities $\widehat{\pi}_I$ of the response. In particular, we are assuming that $P(R_i = 1)$ depends only on the variables O_j , and given these, not on X_i itself. IPW estimates are obtained by calculating weights, which is given as: $\text{Weight}_i = \frac{1}{\widehat{\pi}_I}$. The model of interest is then refitted using the complete cases and the calculated weights by passing the variable containing the estimated weights to the model.

4.8 Double Robustness

Double robustness is a technique that protects the estimates from two possible sources of error. DR produces valid and consistent estimates even if either the outcome model or the weighted model is misspecified. This paper used Inverse Probability Weighted Regression Adjustment (IPWRA) DR method which possesses this double protection property. Using IPWRA, the outcome model was fitted to obtain inverse probability of the weights. The weights estimate missing-data-corrected regression coefficients that are subsequently used to compute the potential outcome means. DR involves both a model for the weights and a predictive model for the missing observations given the observed ones. It was developed as a way of improving efficiency in the results from IPW. The method combines MI and IPW

by including the inverse probability weight $\{P(Y_i \text{is observed}|X_i)\}^{-1}$ to give doubly robust estimators. The process of achieving DR estimators involves specifying the model relating the outcome and covariates of interest, then specifying a model of the probability of observing the variables. Lastly a joint model of the mean of (functions of) partially observed variables given fully observed variables is specified.

5 Results

This analysis was based on 131,508 regular donors who met the inclusion criteria. We observed high amount of missing values in deferral status and number of blood donations variables (Table 1). The initial analysis involved the description of donor characteristics based on donors that had complete information. The median age for donors was 21 years, interquartile range (19, 24 years) and the median number of donations was 3 ranging from 3 to 4. There were 92.2% active donors, 90.4% unmarried donors with 80.7% male donors, and 50% as donors of blood group O.

Donor's marital status was associated with deferral status (P<0.001) where majority of the active donors were not married. There was also an association between donor's sex and deferral status (P=0.009) where 84.1% of the active blood donors were male. We also observed an association between HIV status and deferral status (P<0.001) with more HIV- donors being active donors. There was an association between donor's syphilis status and their deferral status (P<0.001) with majority of active donors being those without the disease (Table 2).

In multivariable logistic regression using MI, we observed an association between deferral status with number of blood donations made and syphilis disease status. Having an increased number of blood donations, was associated with reduced odds of being deferred for donation OR: 0.90 (95% CI: 0.82, 0.98; p=0.02) whereas having syphilis was associated with higher odds of being deferred for donation OR: 1.24 (95% CI: 1.05, 1.48; p=0.01). The findings from MI were consistent with estimates obtained through DR but results obtained from CC and IPW methods were less efficient (Table 3).

Table 1 Percentage of missing values per variable

Variable	Observed	Missing	% Missing
Deferral	4,784	126,724	96.4
HIV status	131,173	335	0.3
Marital status	130,974	534	0.4
Blood donations	5739	125,769	95.6
Syphilis	131,478	30	<0.1

Table 2 Distribution of blood donor deferral status by social-demographic factors and medical history using complete cases

Variable	Active donor n(%)	Deferred donor n(%)	p-value
Married	365 (87.53)	52 (12.47)	<0.001
Not married	4032 (92.69)	318 (7.31)	
Male	3712 (91.79)	332 (8.21)	0.009
Female	700 (94.59)	40 (5.41)	
<20 years	667 (93.03)	50 (6.97)	0.965
20 to 35 years	1358 (93.46)	95 (6.54)	
>35 years	123 (93.89)	8 (6.11)	
2 to 10 donations	4214 (92.09)	362 (7.91)	0.102
>10 donations	198 (95.19)	10 (4.81)	
HIV-	4380 (93.29)	315 (6.71)	<0.001
HIV+	21 (30.00)	49 (70.00)	
Discordant	2 (66.67)	1 (33.33)	
Anemia	2 (100)	0 (0.00)	1
No anemia	4410 (92.22)	372 (7.78)	
Weightloss	1 (100)	0(0.00)	1
No weightloss	4411(92.22)	372(7.78)	
Blood pressure	4(100)	0 (0.00)	1
No Bp	4408 (92.22)	372 (7.78)	
Syphilis	24 (30.00)	56(70.00)	<0.001
No syphilis	4388 (93.28)	316 (6.72)	
Pregnancy	80 (96.39)	3 (3.61)	0.153
No pregnancy	4332 (92.15)	369 (7.85)	

Table 3 Adjusted odds ratio and 95% Confidence Interval (CI)

Variables	CC n=4751	MI	IPCW n=4751	DR n=4751
Intercept	0.05(0.03, 0.07)	0.12 (0.09, 0.17)	0.05 (0.04, 0.08)	3.04 (2.57, 3.60)
Not married	1	1	1	1
Married	1.54(1.07, 2.22)	1.01 (0.92, 1.12)	1.57 (1.07, 2.27)	1.24 (0.91, 1.69)
Female	1	1	1	1
Male	1.61 (1.11, 2.33)	1.02 (0.93, 1.11)	1.51 (1.07, 2.15)	1.02 (1.01, 1.04)
Donations	0.94 (0.89, 0.98)	0.9 (0.82, 0.98)	0.92 (0.86, 0.98)	1.18 (1.12, 1.23)
HIV−	1	1	1	1
HIV+	36.07(21.19, 61.42)	1.1 (0.96, 1.27)	35.92 (21.05, 61.30)	1.43 (0.71, 2.89)
Discordant	6.44 (0.57, 72.97)	1.1 (0.52, 2.34)	5.57 (0.60, 51.40)	0.41 (0.03, 4.87)
Syphilis−	1	1	1	1
Syphilis+	35.06 (21.27, 57.85)	1.24 (1.05, 1.48)	34.86 (21.28, 57.10)	1.46 (0.75, 2.86)

Note: CC: Complete Case, MI: Multiple Imputation, IPCW: Inverse Probability Weighting, DR: Double Robustness

6 Discussion

Blood donation deferral protects both the blood donor and the recipient. Deferral occurs due to various reasons leading to either temporary or permanent deferral. With temporary deferrals, donors return for donation at a later time until they meet donation requirements. This study modelled the factors associated with blood donor deferral. With the high levels of missing values in our data we applied MI, IPW, and DR-IPW methods to appropriately correct for missing values. Through results obtained from MI, we observed that number of blood donations made as well as syphilis disease status was associated with deferral status.

MI produced more efficient and precise estimates as demonstrated by the narrow confidence interval by MI compared to that from CC and IPW (Table 3). The imputation model included the covariates that made MAR assumption plausible but if the imputation model is wrong, MI gives invalid estimates. The plausibility of the MAR assumption ensures correction of bias in parameter estimates. The MI method also accounted for both within and between imputation variations which is in line with the results from a study by Chinomona and Mwambi (2015). The fact that MI imputes the missing values multiple times demonstrates the power of the method to account for the uncertainties in the missing values.

IPW did not eliminate the bias introduced by CC when accounting for the missing values in this study. This is because the method is based on complete cases by assigning the weight to the fully observed variables. Due to the magnitude of the missing values in this study, IPW did not correct for the bias introduced hence the estimates were inefficient despite passing the inverse of the probability of observing the incomplete values given the observed values to the logistic regression model. This is in line with Carpenter et al. (2006) who commented on the difficulty to obtain efficient parameter estimates through IPW.

CC produced inefficient estimates since it discarded observations that were not fully observed leading to a reduction in sample size. The lack of efficiency in the estimates obtained from CC analysis clearly indicates that the fully observed values were not a true representation of the population under study. Chinomona and Mwambi (2015) also found that this method as well as other ad-hoc methods are not principled to account for the randomness in the process of replacing missing values. If the analyses from this study resorted to complete case analyses invalid inferences would have resulted.

MI accounts for both the within and between imputation variability. On the other hand double robustness improves the efficiency of IPW despite the method being applied only on the complete values. The attractiveness of applying this method is its double protection property that guards against model misspecification, for both the data model and the imputation model. Through this method, consistent estimates will still be obtained if the misspecified model is either the model of the probabilities of observing the data or the model of the joint probability of the fully and partial observed values and not necessarily both models. Carpenter et al. (2006) stated that IPW on its own will, however, give inconsistent estimates if the former model is

misspecified, whereas MI imputation gives inconsistent estimates if the later model is misspecified.

7 Limitations

The study used secondary data hence, the researcher had no way of knowing the reasons for missing data, despite considerable efforts to find out. The main missingness mechanism assumption used in these analyses was MAR. This is because literature indicates that MAR is the most plausible assumption in many scenarios (Chinomona & Mwambi, 2015). Based on the MAR assumption and the results obtained from CC, IPW, MI, and DR-IPW, the MI and the DR would be considered as the most reliable results in the researcher's opinion for the analyses from this study. This is based on theory.

For a more accurate analysis of study findings when faced with high proportion of missing values, double robustness methods should be the preferred method to account for the missingness because they are robust to model misspecification. The method, however, is not available in most existing software hence limiting its application. As such, multiple imputation which has shown to produce similar estimates to DR estimates can be used to handle missing data.

8 Conclusion

Based on the variables used for this analysis and results from MI method, regular donors who donated blood several times are less likely to be deferred for donation compared to donors with a reduced number of blood donations made. This can be due to the awareness among the regular donors on blood donation requirements hence the minimal risk of deferral among them.

Missing values affect generalizability of study findings when the analysis is based only on fully observed values. The missing data from this study was based on MAR assumption which is considered to be very plausible in practice. Under the MAR assumption made in this study, MI and DR produced more precise estimates compared to CC and IPW methods as evidenced by narrow confidence intervals from their estimates.

9 Disclosure

The authors declare that they have no competing interests. This work was presented at the International Biometric Society SUSAN conference, Lilongwe, Malawi 22–24 August, 2017, and a poster presentation at Clinic on the Meaningful Modeling

of Epidemiological Data (MMED). Muizenberg, South Africa, May 28 to June 8, 2018.

Authors' contributions	EK	MM
Research concept and design	✓	✓
Data management	✓	
Data analysis and interpretation	✓	✓
Critical revision of article	✓	✓
Final approval of article	✓	✓
Statistical analysis	✓	

Acknowledgments Many thanks to Prof. C. Chasela and MBTS team for making the MBTS dataset available for this analysis. To Dr. Maganizo Chagomerana for his constructive comments.

Appendix

STATA CODE

```
capture log using THESIS_DATA.log, replace
use"C:\Users\ekudowa\Documents\THESIS_DATASETS\
    mbtswithcords1.dta", clear set more off
********keep first observations in each donornumber:::
    ::::::::::::::::::
bysort donorno: gen nkey=_n
keep if nkey==1
**exclusions of some observations ::::::::::::::::::::
    ::::::
gen lastdonationyear=year( lastdonationdate )
keep if lastdonationyear>=2014
keep if number of donations>1
****generating new variables and recoding variables ::
    :::::::
gen birthyear=year( dateofbirth )
gen age=(lastdonationyear-birthyear)
recode age (16/19=1) (20/35=2) (36/100=3), gen(agecat)
recode number of donations (2/10=1) (10/100=2),
    gen(donationscat)
label define agelab 1"<20" 2"20-35" 3">35"
label values agecat agelab
```

```
***********coding marital status to two categories****
    ******
gen married =1 if marital==1
replace married=0 if marital==0|marital==2|marital==3|
    marital==4
label define married 1 "married" 0 "not married"
label values married married
***********coding donor status to two categories******
    *****
gen deferred =0 if status==0
replace deferred=1 if status==1|status==2
label define deferralstatus 1 "yes" 0 "no"
label values deferred deferralstatus
*************descriptive stats before imputation ****
    ******
su age, d
su number of donations, d
ta marital
ta sex
ta status
ta bloodgroup
************for univariate, bivariate and multivariate
    analysis:::
*******crosstabulation for IVs vs DV::::::::::::::::::
    ::::
ta married deferred, chi2 col
ta sex deferred, chi2 col
ta agecat deferred, chi2 row
ta donationscat deferred, chi2 row
ta hiv deferred, exact row
ta anaemia deferred, exact row
ta weightloss deferred, exact row
ta bp deferred, exact row
ta syphilis deferred, chi2 row
ta pregnant deferred, chi2 row
************FITIING UNIVARIATE ANALYSIS MODELS::::::::
    ::::
logit deferred ib0.married, or
logit deferred ib0.sex, or
logit deferred ib3.agecat, or
logit deferred age, or
logit deferred number of donations, or
logit deferred ib2.donationscat, or
logit deferred ib0.hiv, or
logit deferred ib0.syphilis, or
```

```
logit deferred ib1.pregnant, or
*********FITIING A MULTIVARIATE ANALYSIS MODEL
    COMPLETE CASE:
logistic deferred ib0.married ib0.sex number of
    donations ib0.hiv
ib0.syphilis agecat pregnant bp anaemia weightloss
***removing perfect predictors:::::::::::::::::::::::
    ::::::
logistic deferred ib0.married ib0.sex number of
    donations ib0.hiv ib0.syphilis
**********generating dummy variables for missing
    value******
gen deferred_flag=1
replace deferred_flag=0 if deferred==.
gen married_flag=1
replace married_flag=0 if married==.
gen age_flag=1
replace age_flag=0 if age==.
gen noofdonation_flag=1
replace noofdonation_flag=0 if number of donations==.
****performing ttest for missing variables with their
    missing flags
foreach var of varlist deferred_flag -
    noofdonation_flag{ display "`var'"
ttest sex, by(`var')
}
*********missing values table and plot:::::::::::::
    :::::
gen mis=0
replace mis=1 if deferred==.| number of donations==.
label define misv 1"missing" 0"not missing"
label values mis misv
catplot mis agecat ,  percent(agecat) recast(bar) ///
title("Distribution of missing values by age")
    ytitle("%missing") ///
asyvars bar(1, color(blue))bar(2, color(red))
    saving(age)
catplot mis sex ,  percent(sex) recast(bar) ///
title("Distribution of missing values gender")
    ytitle("%missing") ///
asyvars bar(1, color(blue))bar(2, color(red))
    saving(gender)
catplot mis hiv ,  percent(hiv) recast(bar) ///
title("Distribution of missing values by HIV")
    ytitle("%missing") ///
```

```
asyvars bar(1, color(blue))bar(2, color(red))
    saving(hiv)
catplot mis bp ,  percent(bp) recast(bar)  ///
title("Distribution of missing values by bp")
    ytitle("%missing") ///
asyvars bar(1, color(blue))bar(2, color(red))
    saving(bp)
*****missing data summaries::::::::::::::::::::::::::
    ::::::
misstable summ deferred married number of donations
    hiv syphilis
misstable patterns deferred married number of
    donations hiv syphilis
mdesc deferred number of donations married hiv
    syphilis pwcorr deferred sex married number of
    donations hiv syphilis,
    star(0.05) sig
************multiple imputation for the missing values
    :::::
mi set mlong
mi register imputed deferred  number of donations
set seed 888
mi impute chained (logit) deferred (regress) number
    of donations,
    augment add(10)
mi estimate: logit deferred ib0.married ib0.sex ///
number of donations ib0.hiv ib0.syphilis, vce(robust)
mi estimate, or
mi xeq: ta married
mi estimate, vartable
****Inverse Probability Weighting:::::::::::::::::
    :::::::
logit deferred_flag ib0.married ib0.sex  ib0.hiv
    ib0.syphilis, or
predict fitprob1
gen idefprob= 1/fitprob1
hist idefprob
logit noofdonation_flag ib0.married ib0.sex  ib0.hiv
    ib0.syphilis, or
predict fitprob2
gen inoofdprob= 1/fitprob2
hist inoofdprob
logistic deferred  ib0.married ib0.sex number of
    donations ///
ib0.hiv ib0.syphilis [pweight=idefprob]
```

```
    [pweight=inoofdprob]
******generating an indicator for incomplete
    records::::::::
gen imis=1
replace imis=0 if deferred==.|number of donations==.
logit imis ib0.married ib0.sex  ib0.hiv ib0.syphilis,
    or
predict fitprob
gen iprob= 1/fitprob
hist iprob
logistic deferred  ib0.married ib0.sex number of
    donations ///
ib0.hiv ib0.syphilis [pweight=iprob]
gen iprob2=iprob
replace iprob2=60 if iprob2>=60 & iprob2<.
logistic deferred  ib0.married ib0.sex number of
    donations ///
ib0.hiv ib0.syphilis [pweight=iprob2]
**** double robustness using IPRA AND AIPW::::::::::
    :::::::
teffects ipwra (deferred, logit) ( sex number of
    donations ///
ib0.married  ib0.hiv ib0.syphilis, logit ),
    vce(robust) aequations
teffects aipw (deferred, logit) ( sex number of
    donations ib0.married ///
ib0.hiv ib0.syphilis, logit ), vce(robust) aequations
```

References

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*. https://doi.org/10.1016/j.jsp.2009.10.001.

Carpenter, J. R., Kenward, M. G., & Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society*. https://doi.org/10.1111/j.1467-985X.2006.00407.x.

Chinomona, A., & Mwambi, H. (2015). Multiple imputation for non-response when estimating HIV prevalence using survey data. BMC Public Health. https://doi.org/10.1186/s12889-015-2390-1.

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researcher. SpringerPlus. https://doi.org/10.1186/2193-1801-2-222.

Eekhout, I., de Boer, R. M., Twisk, J. W. R., de Vet, H. C. W., & Heymans, M. W. (2012). Missing data: A systematic review of how they are reported and handled. Epidemiology. https://doi.org/10.1097/EDE.0b013e3182576cdb.

Gordon, C. A. (2010). *Investigating statistical approaches to handling missing data in the context of the Gateshead millennium study*. University of Glasgow. Available via DIALOG. https://theses.gla.ac.uk/2312/1/2010gordonmscr..pdf. Cited 20 Jan 2019.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer. Available via DIALOG. https://public.eblib.com/choice/publicfullrecord.aspx?p=3085295. Cited 20 Jan 2019.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*. https://doi.org/10.4097/kjae.2013.64.5.402.

Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials*. Washington DC: National Academics Press.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*. https://doi.org/10.1093/biomet/63.3.581.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *The BMJ*. https://doi.org/10.1136/bmj.b2393.

van Buuren, S. (2012). *Introduction: Concepts of MCAR, MAR and MNAR*. Boca Raton, PL: CRC Press.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v045.i03.

Watanabe, M., & Yamaguchi, K. (2004). *Incomplete data and the generation mechanisms*. New York: Marcel Dekker Inc.

Yuan, Y. (2014). Sensitivity analysis in multiple imputation for missing data. In *Proceedings of the SAS global forum 2014 conference*. Available via DIALOG. https://Support.Sas.Com/Resources/Papers/Proceedings14/SAS270-2014.Pdf. Cited 25 Jan 2019.

Index

A
Absolute Risk Difference (ARD), 5
Absolute standardized differences (ASD), 60–65, 67–71, 74
Accelerated failure-time (AFT) models, 22, 288–291, 296
Addis Abeba city administration, 435, 443, 447
Ad-hoc methods, 460, 469
Afar region, 26, 34, 435, 436, 449
Amhara region, 26, 435
Ammag distribution, 289
Antenatal care (ANC), vii, 57, 58, 262–265, 267–280
Antiretroviral drugs, vii, 205, 214–221
Association, viii, ix, 4, 12, 17, 22, 42, 43, 45, 57, 58, 60, 75, 76, 79, 105, 121, 146, 175, 176, 236, 237, 241, 248, 251–256, 262, 267, 277, 278, 300, 302, 321, 323, 341, 348, 357, 364, 367–369, 415–417, 420, 429–431, 449, 467
Average treatment effect, 48, 49, 52, 65, 71, 80, 98

B
Baseline hazards, 288, 320, 326, 380, 391, 397, 401, 416, 442
Baseline levels, 14, 34, 330
Bayes factor, 299, 300
Bayesian framework, 22, 108, 297, 299, 302, 321, 323, 370
Bayesian inference, viii, 297–300, 312, 341, 346
Bayesian information criterion (BIC), 420, 423, 428, 429
Benishangul–Gumuz region, 36, 435, 446
Biased estimates, 42, 378, 405, 430, 460
Binary outcome, 14, 55, 56, 79, 148, 152, 170, 206, 358, 363
Biomedical study, 205, 414, 430, 431
Birth cohort, 13, 14, 17, 434, 436, 443–446
Birth outcome, 22
Bivariate mixture models, 268
Block sampling, 298–299
Blood donation, 460, 463, 464, 467, 469, 470

C
Causal inference, vi, 45, 48, 81, 82
CD4 count, vii, 120, 205, 213, 214, 216, 217, 219, 414
Censoring, 289, 297, 298, 325, 380, 386, 390, 405, 415, 416, 436, 440, 442, 461
Child growth, 44, 56, 58, 60, 75–76, 79, 81, 82
Child growth measurements, 81
Childlessness, 26
Child malnutrition, 44
Child nutritional outcomes, 44–45, 57, 58, 60–65, 72, 73, 75, 77, 78, 82
Children ever born, 22, 26, 27, 30, 31
Clayton copula, 265
Clustered data, 4, 378, 392, 434
Clustering, vii, ix, 4, 81, 82, 99–101, 235–258, 268, 340, 380, 434–438, 440–443, 446, 449

D.-G. (Din) Chen et al. (eds.), *Modern Biostatistical Methods for Evidence-Based Global Health Research*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-11012-2

Cohen's d, 171, 189
Competing risks, viii, 413–431
Complementary feeding, 44, 57–60, 62, 64–66, 68, 70, 71, 74–80, 82
Complete case (CC), 390, 460, 461, 464, 466, 468, 469
Conditional autoregressive (CAR), viii, 266, 339–352
Conditional logistic regression, 49, 56, 75
Confounder balance, 54–56, 71, 79, 81
Confounder bias, 42–56, 76, 80, 81
Confounders, 42, 43, 47, 57, 58, 60, 61, 63, 65, 67, 69, 71, 74, 76, 80, 81
Consistency, 45, 46, 51, 109–111
Consistent estimates, 150, 466, 469, 470
Continuous outcome, 65, 71, 76, 126, 140, 170–171, 189–190, 192, 195, 206, 278, 419
Control group, 42, 43, 47, 49–51, 54, 99, 100, 148, 149, 156, 160, 171, 175, 184–186, 214, 216
Copula-based regression, 268
Correlated survival data, 378
Couples' education, 11
Covariance, viii, 206–214, 220, 221, 241, 264, 298, 341, 361, 380, 393, 415, 417, 418
Covariance structures, 206, 207, 210–214, 221
Covariate, 13, 23, 43, 107, 191, 206, 245, 265, 288, 320, 341, 358, 380, 414, 434, 466
Cox-specific model, 392, 404, 414, 415, 430, 446, 447

D

Data, 2, 10, 22, 42, 93, 137, 168, 205, 235, 263, 288, 320, 341, 356, 378, 414, 434, 460
Demographic and Health Surveys (DHS), vi, vii, 10, 17, 22, 23, 34, 56, 58, 59, 267, 268, 287–316, 322, 341, 342, 390, 434–438
Dersimonian-Laird estimator, 149, 153, 187, 191
Design matrix, 207, 360, 395
Developing countries, 10, 22, 262
Deviance information criterion (DIC), 345, 347–348, 351
Diagonal reference models (DRM), v, vi, 10, 12–18
Dire Dawa city administration, 26, 34, 446, 447
Discount factor, 327
Double robustness, ix, 459–475
Dynamic modelling, 321, 328, 330, 332

E

Education, 10, 23, 57, 122, 196, 269, 301, 320, 348, 357, 390, 434
Educational gradients in family initiation, 319–335
Educational gradients of parity progression and childlessness, 21–39
Efficacy, 12, 140, 156, 414, 419
Efficient estimates, 460, 469
Entry into marriage, viii, 289, 300, 309, 320, 328, 330, 333
Ergodic mean theorem, 299
Eritrea, v, vi, 9–19, 321, 333, 435
Error, 14, 15, 23, 37, 71, 91, 98–102, 110, 138–142, 144, 145, 147, 150, 153, 154, 162, 163, 192, 214–217, 267, 273, 279, 280, 290, 291, 361, 378, 382, 393, 403, 404, 416, 417, 430, 441, 466
Estimation, vii, 25, 42, 45, 47–48, 71, 76, 80, 141, 149, 150, 152, 154, 162, 163, 191–193, 205, 208–210, 217, 221, 243–246, 258, 289, 343, 346, 370, 378, 380–381, 393, 395, 403, 418, 434, 435, 440, 447, 460
Ethiopia, vi, ix, 21–39, 433–457
Event, 100, 125, 153, 180, 185, 186, 265, 288–290, 292, 300–302, 306, 308, 310, 320, 322, 325, 326, 328, 333, 378, 380, 382, 383, 390, 394, 396–398, 414–418, 420, 421, 425–428, 430, 435, 436, 439–442, 443, 449
Exchangeability, 46
Exclusive breastfeeding (EBF), 44, 45, 57–62, 65–68, 71–76, 79, 82
Exponential distribution, 292, 326, 332
Exposure time, 325, 326, 436
Extended generalized gamma (EGG) model, vii, 287–316

F

Family planning, 22
Family size, 36
Fertility, vi, 22, 23, 26, 36, 435
Fixed-effects, vi, 97–99, 105, 138–143, 146–148, 156–159, 163, 182, 184, 186, 188, 191, 196, 198, 207, 209, 210, 346, 380, 386, 388, 389, 418, 441
Fixed-effects model, 98, 141–143, 147, 148, 157, 182, 184
Follow-up, 105, 123, 124, 126, 190, 263, 300, 321, 378, 398, 414, 419, 420, 423

Forest plot, 96, 101, 102, 110–112, 116, 118, 119, 122, 158, 183, 186–189, 193, 219, 220
Frailty models, 380, 390, 392, 435, 440–441, 443–447, 449
Framework, viii, 22, 25, 57–58, 91, 95, 108, 127, 139, 144, 240, 268, 278, 288, 291, 297, 299, 302, 321, 323, 325, 332, 370, 415, 429, 430, 435
Frank copula, 265, 271
Fully observed, 461, 466, 467, 469, 470
Funnel plot, 125, 183, 187, 188, 191, 192
Funnel plot asymmetry, 191, 192

G
Gambela Region, 27, 28, 33–35, 36, 435, 437, 438, 445, 446, 448
Gamma distribution, 243, 292–294, 296, 297, 442
Gaussian copula, 266, 267
Gaussian Markov random field (GMRF), 268, 269, 343, 351
Generalised joint regression model (GJRM), 268, 278
Generalized additive models, 321
Generalized linear models (GLMs), viii, 25, 152, 162, 250, 258, 265, 266, 341, 342, 358, 363, 369, 370, 393
General linear mixed model, 205, 206
Geographically weighted regression (GWR), 341, 343
Glass' delta, 171, 189

H
Half-normal distributions, 289
Harari Region, 27, 28, 33–35, 435, 437, 438, 445, 446, 448
Hazard/intensity functions, 289–291, 294, 320, 321, 325, 326, 380, 383, 416, 439–441, 442
Health care utilization, vii, 10–19, 262–280, 300, 321
Health care utilization Eritrea, v, 10–19
Health decision, vi, 92, 93
Hedge's g, 171, 189
Hemoglobin level, 414, 419–423, 424, 426–428, 429
Heterogeneity, vi, 32, 56, 98, 99, 101–105, 110, 111, 114–116, 117, 119–121, 125, 138, 140, 142–146, 147, 149, 152,156, 158–160, 162, 163, 167, 168, 178, 182–184, 186, 188–196, 198, 217, 238, 434, 447
Hierarchical, 81, 82, 108, 117, 139, 149, 152, 343, 434, 447
HIV studies, 205, 221
Hospital delivery, 11–12, 14–17
Households, ix, 4, 22, 26, 29, 34, 56, 57, 60, 269, 270, 274, 276, 342, 390, 391, 434–438, 440–443, 447, 449
Human immunodeficiency virus (HIV), vi, vii, viii, 3–5, 57, 58, 60, 61, 63, 67, 69, 74, 75, 120, 170, 176, 177, 179, 214–221, 340–352, 378, 463, 464, 467, 468
Hunter-Schmidt estimator, 191

I
I^2, 159
Ideal number of children, 17, 22
Ignorability assumption, 45, 46, 48, 51
Imputations, ix, viii, 79, 80, 214, 429, 460–475
Incomplete values, 469
Independent effect sizes, 205, 211–213
Inefficiency factor (IF), 299
Inefficient estimates, 469
Infectious disease, 92
Integrated nested Laplace approximation (INLA), 341, 346, 352
Intended number of children, 17
Intermittent missingness, 461
Intervention studies, 113–115, 118, 120, 168, 170, 171, 183–190, 418
Inverse ammag distribution, 289
Inverse exponential distribution, 294
Inverse gamma distribution, 294, 305, 309
Inverse-gamma priors, 327
Inverse probability, 43, 71, 466
Inverse probability weighting (IPW), ix, 51–54, 76, 460–475
Inverse Rayleigh distribution, 289, 294
Inverse variance method, 98, 168, 175, 187, 192, 194, 195
Isotropic, 264

J
Joint distribution, 207, 264, 345
Joint models, vii, viii, 266–268, 274, 276, 277, 414–431, 467

L
Likelihood, 25, 53, 114, 115, 117, 139, 163, 208, 248, 267, 289, 297–299, 321, 325–327, 328, 344, 348, 358, 361, 362, 369, 370, 380, 393, 439, 440, 442, 443, 460, 461
Logistic regression, ix, viii, 10, 14–16, 17, 23, 24, 36, 47–49, 52, 65, 75, 80, 153, 162, 163, 333, 346, 355–371, 464, 466, 467, 469
Log-normal distribution, 291, 305, 441, 443, 449
Log-normal priors, 327
Log predictive density score (LPDS), 289, 300, 305–309, 310
Longitudinal, vii, viii, 3, 99, 205, 206, 211–213, 214, 219–221, 414–431, 461
Longitudinal data, ix, viii, 4, 81, 414, 415, 429, 430
Longitudinal meta-analysis, vi, vii, 204–228
Longitudinal multiple effect sizes, 205

M
Malawi DHS, vii, 57, 61–69, 70, 72–74, 76–80, 267, 269, 272, 390, 392, 405
Mantel–Haenszel test**,** 55, 56, 76, 179, 180
Markov chain Monte Carlo (MCMC), 25, 289, 298, 299, 321, 327, 346, 362, 363, 366, 370, 465
Maternal health, vi, 5, 10, 22, 57, 81, 262, 267, 271, 277, 278
Maternal HIV, vi, 57, 58, 60, 63, 69, 74, 75
Maximum likelihood (ML) estimator, viii, 48, 191, 381, 392, 393
McNemar's test, 49, 55, 56, 65
Mean difference (MD), 54, 97, 107, 140, 150, 168, 171, 189–191, 194, 196, 198, 216, 219, 220
Meta-analysis, vi, vii, 90–128, 137–164, 168–199, 204–228
Meta-analysis model, 109, 114, 139, 141, 142, 159, 163, 164, 205–208
Metabias, 192
Metabin, 186
Metacont, 189–191, 193, 198
Meta-regression, vi, 104–105, 107, 111, 143, 146–147, 155, 158–160, 161, 163, 192, 197–198
Metropolis–Hasting algorithm**,** 298–299, 310
Missing data, ix, viii, 45, 103, 105, 415, 460–466, 470
Missingness mechanism, 462, 464, 466, 470
Missing values, 460–462, 464–466, 467, 469, 470
Mixed-effects, viii, 149, 152, 162, 181, 381, 382, 393, 397, 398, 415, 417, 419, 423, 424, 430
Model comparison, 299–300, 347–348, 429
Model evaluation, viii
Modelling/modeling**,** v, ix, vii, viii, 4, 5, 10–19, 117, 139, 149, 152, 163, 257, 262–280, 320–335, 363, 378, 414–431, 434–457
Model misspecification, 48, 469, 470
Monotonic missingness, 461
Mother's education, 26, 30, 31, 60, 61, 67, 74, 274, 277, 390, 391
Multilevel modelling, ix, 363, 434–457
Multiple imputation, ix, 79, 80, 460–475
Multiple outcomes, vi, 81, 205, 206
Multivariate distribution, 264, 265, 465
Multivariate effect sizes, 205
Multivariate Imputation by Chained Equations (MICE) algorithm, 465
Multivariate meta-analysis, vi, 108, 205, 211, 212, 217
Multivariate normal distribution, 207

N
Nationalities and Peoples Region, 435
Nigeria, 339–341, 351, 356, 357, 367
Non-ignorable, 462
Non-linear effects, 266, 384
Non-linear models, 393
Non-random missingness, 42, 43, 123
Non-stationarity, 341
Nutritional interventions, 60–65

O
Observational studies, 41–83
Observed values, 147, 237, 242, 462, 469, 470
Odds ratio (ORl), 14–16, 55, 76, 80, 101, 110, 114, 115, 121, 140, 153, 158, 160, 163, 171–173, 176, 180–183, 214, 216, 217, 359
Oromia Region, 436, 446

P
Package, 56, 99, 109, 119, 169, 198, 278, 420
Paired data, 43, 49, 110, 114, 268
Parameter, 13, 14, 25, 32, 148–150, 170, 194, 208–210, 220, 243, 259, 264–267, 288, 289, 291–293, 297, 305, 309, 327,

343–345, 381, 386, 397, 403, 415, 418, 439, 442
Parametric inference, 288
Parametric survival models, 289
Parasite count, 419, 420, 422–430
Parity, 22, 31, 34
Parity progression, 21–39
Partial Likelihood, 380, 393, 439, 440, 443
Particle Filter (PF), 321, 333
Partners education, 10–12, 15–17
Patients, 99, 100, 107, 116, 120, 126, 138, 147, 152, 161, 171–173, 196, 214, 220, 363, 370, 378, 419, 423, 427, 430
Paule-Mendel estimator, 191
Penalized likelihood, 443
Peto odds ratio, 180–183
Piece-wise exponential model, 321, 326
Polynomial splines, 321
Pooled estimates, 178, 184, 195, 465
Positivity, 46, 116, 118
Posterior distribution, 245, 297–299, 310, 321, 327, 333, 362, 366, 367, 370, 418, 465
Predicting, 423, 426–430
Prenatal care, 10
Prior distribution, 245, 343–345, 362, 363
Probit model, 21–39, 48
Profile, 15, 127, 209, 419, 422
Propensity, 10, 17, 22, 31, 32, 36
Propensity score, 41–83
Propensity score matching (PSM), 43, 48–51, 65, 76, 80–82
Proportional hazards models, 288, 320, 321

Q
Q test, 115, 198

R
Random effects model, 98, 99, 117, 142, 184, 198, 267
Randomized controlled trials, 205
Random missingness, 460, 465
Randomness, 239, 240, 469
Random time effects, 205, 211–214, 217, 219, 221
Random walk, 298, 310, 321, 327, 342, 343
Rayleigh distribution, 289, 294
Reciprocal Weibull distribution, 292, 305
Region, 2, 5, 14, 26, 34, 105, 117, 118, 177–179, 197, 272, 274–276, 340, 343, 346, 350, 364, 435, 436, 449
Regional differences, 26, 433–457
Regional differences, Household random effects, 438
Relative risk (RR), 168, 171, 416, 417
Research, 1–6, 36, 76, 90–94, 125, 139, 205, 257, 351, 357, 369, 429
Residence, 10, 12, 13, 57, 60, 269, 274, 276, 301, 309, 322, 333, 346, 391, 434, 444
Residential gradients in family initiation, 319–335
Residuals, 204, 206, 207, 221, 249, 382, 392, 405, 429
Restricted maximum likelihood (REML), 208, 210
Restricted Maximum Likelihood estimator, 191
Retrospective Surveys, 333, 463
Right-censored data, 298
Risk ratio, 97, 103, 115, 140, 171, 173–175, 180, 186, 196, 204
R software, 140, 168, 169, 191, 345, 363
Rubin rule, 465

S
Selection bias, 10, 42, 54
Semi-parametric inference, 439
Semi-parametric survival models, 321
Sensitivity analysis, 55–57, 76, 80, 105, 195–196, 462
Sequential decision process, 23
Sequential models, 25
Sequential Monte Carlo (SMC), 328, 333
Severe malaria, 97, 413, 420, 423, 426–428, 430, 431
Shared frailty, 435, 438, 441, 447, 449
Simultaneously and conditionally autoregressive, 344
Smoothing, 242, 247, 262, 269, 327
Social mobility, 10, 12
Socio-demographic, 289, 350, 357, 363, 367, 369, 439
Somali region, 26, 435
Southern Nations, 435
Southern Nations, Nationalities and Peoples Region (SNNPR), 435, 445
Spatial dependency, 238, 253, 258, 267, 268
Spatial effect, 269, 321, 343, 350, 351
Spatially varying coefficients, 341, 347, 348
Spatial regression model, 268
Spatial statistics, 235
Spatial variation, 275–280
Standard and generalized gamma distributions, 287–316

Standardized mean difference, 140, 168, 171, 190
Stationarity, 341, 344, 351, 366, 367, 370
Stratified modelling, 433–457
Structured, 122, 127, 267, 343, 345, 346, 350
Student-*t* copula, 264, 265, 272
Study bias, 191
Stunting, 44, 56, 58, 75, 76, 79–81
Subgroup analysis, 102–105, 195, 196
Sub-Saharan Africa, 1–6, 81, 92, 351, 420
Survival function, 289, 290, 294, 296, 302–304, 320, 325, 326, 443
Survival models, 22, 289, 321, 332, 333, 378, 380, 382, 384, 392, 395, 397, 403, 405, 414, 434
Survival package in R, 443

T
Tau, 187
Tau^2, 187
The DHS program, 22
Tigray region, 26, 34, 436, 445
Time, 11, 22, 26, 96, 106, 211–213, 251–256, 265, 277, 288, 290, 301, 320, 325, 326, 328–332, 357, 378, 391, 403, 414, 416, 420, 436
Time points, 205, 206, 210–214, 253, 383, 416
Time-to-event data, 320
Timing of first birth, 22, 436
Transition to parenthood, 34, 434, 436, 441–447
Treated group, 49–51
Treatment, 42–46, 48–50, 52, 54, 55, 71, 98–100, 105, 107–114, 120, 121, 138–140, 148, 152, 153, 161, 171, 184, 205, 216, 370, 378, 420, 422

U
Unbiased estimates, 210, 460, 462
Underweight, 44, 56, 58, 60, 65, 71, 75, 76, 79, 81
Univariate meta-analysis, 204, 205
Univariatemissingness, 461
Unobserved heterogeneity, 56, 434, 447
Unstructured, 268, 343, 345, 346, 351

V
Variance-covariance matrix, 49, 109, 207, 212–214, 418
Viral load suppression, 205, 213, 217, 219, 221

W
Wasting, 44, 56, 58, 65, 71, 75, 76, 79, 81
Weibull distribution, 292, 305
Wishart priors, 327
Withdrawal, 419–421, 423, 426–431
Within-study effect sizes, 213–214